Cases of Assessment in Mathematics Education

An ICMI Study

Edited by

Mogens Niss

Kluwer Academic Publishers

[illegible]

Cases of Assessment in Mathematics Education

Published under the auspices of The International Commission on Mathematical Instruction under the general editorship of

Miguel de Guzmán, President Mogens Niss, Secretary

Cases of Assessment in Mathematics Education

An ICMI Study

edited by

Mogens Niss

Roskilde University, Denmark

Kluwer Academic Publishers

Dordrecht / Boston / London

Library of Congress Cataloging-in-Publication Data

Cases of assessment in mathematics education : an ICMI study / edited by Mogens Niss.
p. cm.
Based on an ICMI conference held at Cap Roig, Calonge, Spain, Apr. 11-16, 1991.
Includes bibliographical references.
ISBN 0-7923-2089-1 (alk. paper)
1. Mathematics--Study and teaching--Evaluation--Case studies.
I. Niss, Mogens. II. International Commission on Mathematical Instruction.
QA11.C3214 1993
510'.71--dc20 92-41538

ISBN 0-7923-2089-1

Published by Kluwer Academic Publishers,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Kluwer Academic Publishers incorporates
the publishing programmes of
D. Reidel, Martinus Nijhoff, Dr W. Junk and MTP Press.

Sold and distributed in the U.S.A. and Canada
by Kluwer Academic Publishers,
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers Group,
P.O. Box 322, 3300 AH Dordrecht, The Netherlands.

Printed on acid-free paper

Printed in the Netherlands

TABLE OF CONTENTS

MOGENS NISS

INTRODUCTION

The present book, *Cases of Assessment in Mathematics Education*, is one of two studies resulting from an ICMI Study Conference on **Assessment in Mathematics Education and Its Effects**. The book which is published in the series of **ICMI Studies** under the general editorship of the President and Secretary of ICMI is closely related to another study resulting from the same conference: *Investigations into Assessment in Mathematics Education* (Niss, 1992). The two books, although originating from the same sources and having the same editor, emphasize different aspects of assessment in mathematics education and can be read independently of one another. While the present book is devoted to presenting and discussing cases of assessment that are actually implemented, the other study attempts to critically analyze general and principal aspects of assessment. Naturally, the content of either book is enriched by the materials and perspectives provided by the other one.

In order to put this book and its background into context, the nature and scope of the ICMI studies are outlined briefly below.

Since 1986 the *International Commission on Mathematical Instruction (ICMI)* has been engaged in publishing a series of studies on essential topics and key issues in mathematics education. Previously, the following studies have been published (all by Cambridge University Press): *School Mathematics in the 1990s* (1986), *The Influence of Computers and Informatics on Mathematics and Its Teaching* (1986), *Mathematics as a Service Subject* (1988), *The Popularization of Mathematics* (1990), *Mathematics and Cognition: A Research Synthesis by the International Group for the Psychology of Mathematics Education* (1990).

Depending on the theme under consideration a study may either be *research* oriented or *action* oriented (or both). In either case the aim is to provide an up-to-date presentation and analysis of the state-of-the-art concerning a theme, whether by identifying and describing current research contributions and their findings, or by identifying and discussing crucial, non-rhetorical issues involving genuine controversies or dilemmas and the different positions towards them held by various mathematics educators.

In order to provide a platform for producing an ICMI study the following normal procedure has been adopted (the exception is the study on cognition). The Executive Committee of ICMI appoints a fairly small, international *Program Committee*. Its first task is to write a so-called

M. Niss (ed.), Cases of Assessment in Mathematics Education, 1–7.

Discussion Document that outlines the theme, the aims, and the scope of the study, and presents the items and issues to be dealt with. The Discussion Document is published in international journals (including the official organ of ICMI, *L'Enseignement mathématique*) and newsletters with an invitation to mathematics educators to respond to the Document and to apply for participation in a so-called *Study Conference*.

The Study Conference is held with a limited number (50–100) of individuals and constitutes a working forum of experts and novices with ideas, experiences and expertise to investigate the theme of the study. This investigation is guided by the Discussion Document, assisted by working papers (written by participants), presentations, debates, and group work.

Finally, the study proper is produced and published under the general editorship of the President and the Secretary of ICMI, and based on the written materials and the work done at the Study Conference. As every study is written and edited as an independent publication for a wide international readership, its nature is *not* that of a conference proceedings.

In May 1989 the Executive Committee of ICMI appointed the following international *Program Committee*:

Claudi Alsina, local organizer, Universitat Politècnica de Catalunya, Barcelona, Spain;
Desmond Broomes, University of the West Indies, Bridgetown, Barbados;
Hugh Burkhardt, Shell Centre for Mathematical Education, University of Nottingham, UK;
Mogens Niss, chairman of the Program Committee, Roskilde University, Denmark;
Thomas A. Romberg, National Center for Research in Mathematical Sciences Education, University of Wisconsin-Madison, USA;
David Robitaille, University of British Columbia, Vancouver, Canada;
Júlianna Szendrei, O.P.I. (National Institute of Education), Budapest, Hungary.

The Discussion Document was officially published in *L'Enseignement mathématique*, **36**, fasc.1–2, Janvier-juin 1990, 197–206, as well as in a number of other journals and newsletters.

The *Study Conference* which was held at Cap Roig, Calonge, Spain, 11–16 April 1991, had 80 contributing participants from 25 different countries in Europe, North America and the Caribbean, Asia, Oceania, Africa and the Middle East.

A Note on Terminology

Some terminological clarification may be in order. The field we shall be dealing with frequently uses terms such as assessment, evaluation, tests, exams. However, these words and their counterparts in other languages

carry quite different connotations within different educational systems and contexts. The variation is so large that the same word often has different meanings to different people. We shall confine ourselves to making one distinction, namely between *assessment* and *evaluation*.

While assessment and evaluation are often used interchangeably we shall adopt, as was suggested in the Discussion Document for the present study, the following terminological convention:

Assessment in mathematics education is taken to concern the judging of the mathematical capability, performance, and achievement — all three notions to be taken in their broadest sense — of **students** whether as individuals or in groups, with the notion of "student" ranging from Kindergarten pupils to Ph.D. students. Assessment thus addresses the outcome of mathematics teaching at the *student level*. **Evaluation** in mathematics education is taken to be the judging of **educational or instructional systems**, in its entirety or in parts, as far as mathematics teaching is concerned. Evaluation may concern system components such as curricula, programs, teachers, teacher training, and specific segments of the educational system such as schools or school districts etc. So, evaluation addresses mathematics education at the *systems level*.

When tests and exams are considered to be ways of judging student performance they are special forms of assessment and are thus subsumed under the assessment category. As a contradistinction, when tests and exams are viewed as being part of the modes of operation of an educational (sub)system, or when the outcomes of tests and exams are used as indicators of the quality of such a system, as is the case with international performance comparisons, exams and tests belong to the realm of evaluation. This duality shows features of the general relationship between assessment and evaluation: Assessment items — in particular assessment results, but also assessment modes — may be involved in the judging of system aspects, hence they would form part of an evaluation activity. The converse normally will not hold for evaluation; for instance the appraisal of teachers will often involve a multitude of components having nothing to do with assessment of students. So, the relationship between assessment and evaluation is not a symmetrical one.

In the present study the emphasis will be on *assessment* as defined above rather than on evaluation. Due to the duality just mentioned this does not imply that evaluation issues will not be considered. However, only those aspects of evaluation which have to do with assessment of students will be given explicit attention.

Why a Study on Assessment?

In recent years, assessment has attracted increased attention from the international mathematics education community. There are numerous reasons for this. One seems to predominate. During the last couple of

decades, the field of mathematics education has developed considerably in the area of ideals and goals, and theory and practice, whereas assessment concepts and practices have not developed so much.

The mathematics curriculum has claimed new territory. First, when it comes to *content*, aspects of applications and modeling, cooperation with other subjects on topics of common interest, philosophy and history of mathematics, problem-oriented creativity, explorations and experiments aided by computers and informatics have been included in quite a few programs and curricula round the world. Secondly, we have witnessed a remarkable expansion of the spectrum of *working forms* and student *activities*. Extended investigations of pure and applied mathematics, project work, scientific enquiry and debate, out-of-classroom activities, experimentation, group work, and so forth, are no longer utopian entities in mathematics teaching. As a result, a much broader notion of mathematics and mathematics education has emerged.

These developments have *not*, however, been matched by parallel developments in assessment, where values, notion, and theory, practice, modes, and procedures are concerned. Consequently, an increasing mismatch and tension between the state of mathematics education and current assessment practices are materializing. It may well be the case that the ideals and goals of mathematics education were *never* really in accordance with the assessment modes available to mathematics educators but, as in former times post-elementary mathematics education was offered only to a minority of children and youth, the problems created by the mismatch were, perhaps, less serious, or at least thought by mathematics educators to be less serious. At any rate, expanding the notions of mathematics and mathematics education has undoubtedly widened the gap between contemporary mathematics teaching and traditional assessment practices.

This gap has put assessment on the agendas of mathematics educators. In the interest of truth, it should be said that this is a rather new phenomenon. The development of mathematics teaching during the last three or four decades has emphasized curriculum reform — of different and sometimes even contradictory types, that is true — as the most important task. Concurrently mathematics education, as an academic field, has focused attention on the conditions for and processes involved in the learning of mathematics, in particular regarding the formation and acquisition of mathematical concepts. This largely left assessment out. Thus, it was viewed as a less important factor in mathematics education, a factor that in addition was "external" to mathematics education in several respects. To the extent assessment has attracted the attention of mathematics educators, it has often been due to uneasiness about its role and function. Traditional assessment modes, especially examinations and tests administered "from outside", have, in many cases, formed one of the factors that hindered or slowed down curriculum reform.

Now that curriculum reform *has* been, or is being carried out in many places, the situation has changed. The roles, functions, and effects of assessment in mathematics education should no longer be neglected; rather, they should become objects of investigation and examination for several reasons (see Commission Internationale de l'Enseignement Mathématique, 1990):

- The roles, functions, and effects of contemporary modes of assessment are neither clear, nor well understood.
- Current assessment modes and practices involve conflicting interests, divergent aims, and unintended or undesired side-effects. In particular, it is difficult to devise assessment modes which at the same time: (a) allow us to assess, in a valid and reliable way, the knowledge, insights, abilities, and skills related to the understanding and mastering of mathematics in its essential aspects; (b) provide genuine assistance to the individual learner in monitoring and improving his or her acquisition of mathematical insight and power; (c) help the individual teacher in monitoring and improving his or her teaching, guidance, supervision, and counseling; (d) assist curriculum planners and authorities, textbook authors, and in-service teacher trainers in adequately shaping the framework for mathematics instruction.
- The difficulties involved in devising and employing effective, harmonious assessment modes, free from serious internal and external problems, seem to be fundamental and universal in nature, and hence worthy of being dealt with from an international perspective.

The Content of this Book

It is the purpose of this study to explore selected *cases of assessment in mathematics education*. This is done by presenting and examining current assessment practices in a number of nations, and by identifying and discussing examples, practices, and ideas that will contribute to linking together assessment with the purposes and goals, the implementation and the outcomes of mathematics teaching.

Some chapters give an overall presentation and examination of the state of assessment in mathematics education in countries which have adopted assessment concepts and practices that in some way are "archetypical", i.e. represent rather characteristic formats of assessment recognizable, though not necessarily widespread, in several other countries as well. This is the case with the chapters written by *Luis Rico* (Spain), *Desmond Broomes & James Halliday* (Barbados), *Murad Jurdak* (the Arab countries), *John Dossey & Jane Swafford* (the United States), *Margaret Brown* (the United Kingdom), *Luciana Bazzini* (Italy), and *Gunnar Gjone* (Norway).

Other chapters present national, centralized assessment practices that are unconventional but interesting blends of well-known formats. The chapters by *Hans Nygaard Jensen*, and *Kirsten Hermann & Bent Hirsberg* (Denmark), *Wim Kleijne & Henk Schuring* (the Netherlands), and *Max Stephens & Robert Money* (Australia) are of this type.

A third category of chapters presents the assessment modes adopted for specific curriculum programs or projects that are non-compulsory in their country. The papers in this category include those written by *Edward Silver & Suzanne Lane* (the United States), *Chris Little* (the United Kingdom), *Leonor Cunha Leal & Paulo Abrantes* (Portugal), *Wei Chao-qun & Zhang Hui* and *Cheng Zemin & Lü Shaozheng* (China).

The book is concluded with a chapter (by *Ruth Sweetnam*, the International Baccalaureate, UK) that describes the assessment practices of the International Baccalaureate which is attracting a considerable and increasing number of students from all parts of the world.

When it comes to instances of *innovative or experimental assessment* it should be noticed that these may not only be found with non-traditional curricula but also with centralized, national assessment systems. So, a reader who is particularly interested in the innovative aspects of assessment would be well advised to look for these in all the chapters of this study.

Acknowledgements

A great many individuals have played important roles in creating this study. First of all Claudi Alsina (Universitat Politecnica de Catalunya, Barcelona, Spain) who was the protagonists in organizing the ICMI Study Conference on Assessment in Mathematics Education and Its Effects, and the members of the international Program Committee, many of whom gave very valuable advice in the early stages of the editorial process. It is important to underline that even if only a minority of the participants in the Study Conference have contributed to the two studies resulting from it, the participants' work has exerted an important influence on the editing, the format, and the content of both books. The Editor therefore wants to express his gratitude to all the conference participants.

Special thanks go to Thomas A. Romberg and Joan Daniels Pedro, the National Center for Research in Mathematical Sciences Education, University of Wisconsin-Madison, USA, for their assistance with the final polishing of the papers. In particular, the Editor is deeply grateful to Joan Daniels Pedro who did a magnificent job in polishing and harmonizing the English of the articles and homogenizing their formats. However, any remaining linguistic or editorial flaws are solely the responsibility of the Editor.

Many thanks are further due to Kluwer Academic Publishers, in particular to Peter de Liefde and Nicola Berridge, who have been most cooperative and flexible in the whole process of preparing this book. Last,

but not least, the Editor wants to thank his son Henning Niss who was in charge of producing the camera-ready manuscript.

NOTE

A substantially enlarged version of this introduction is contained in the other ICMI Study on assessment in mathematics education, *Investigations into Assessment in Mathematics Education: An ICMI Study* (Niss, 1992).

REFERENCES

Churchhouse, R.F. et al. (Eds.): 1986, *The Influence of Computers and Informatics on Mathematics and Its Teaching*, ICMI Study Series, Cambridge University Press, Cambridge.

Commission Internationale de l'Enseignement Mathématique: 1990, 'Assessment in Mathematics Education and Its Effects: Discussion Document', *L'Enseignement mathématique*, 2e série, 36, fasc.1–2, Janvier-juin 1990, 197–206.

Howson, A.G., Kahane, J.-P., Lauginie, P., & de Turckheim, E. (Eds.): 1988, *Mathematics as Service Subject*, ICMI Study Series, Cambridge University Press, Cambridge.

Howson, A.G. & Kahane, J.-P. (Eds.): 1990, *Popularization of Mathematics*, ICMI Study Series, Cambridge University Press, Cambridge.

Howson, A.G. & Wilson, B. (Eds.): 1986, *School Mathematics in the 1990s*, ICMI Study Series Cambridge University Press, Cambridge.

Nesher, P. & Kilpatrick, J. (Eds.): 1990, *Mathematics and Cognition: A Research Synthesis by the International Group for the Psychology of Mathematics Education*, ICMI Study Series, Cambridge University Press, Cambridge.

Niss, M. (Ed.): 1992, *Investigations into Assessment in Mathematics Education: An ICMI Study*, Kluwer Academic Publishers, Dordrecht.

Mogens Niss
IMFUFA,
Roskilde University
Denmark

LUIS RICO

MATHEMATICS ASSESSMENT IN THE SPANISH EDUCATIONAL SYSTEM

1. INTRODUCTION

The Spanish form, indeed culture, of assessment is inevitably a product of Spanish society itself, where many habits and customs have undergone great changes in the past few years. To understand the present situation it is helpful to know something of its immediate antecedents.

The Spanish educational system that has been in operation for the past few years is the result of the *General Education Law* of 1970 (Ministerio de Educacion y Ciencia, 1972), which represented a great effort, at the time, of rationalization and modernization of the educational structures which had been formed during the dictatorship.

At the end of the 1960s it was necessary to support and complement the social and economic changes which had taken place in society in general by making equivalent changes in the educational field. The present education system was designed with help and advice from the UNESCO and the OECD. It established a single period of compulsory education, from 6 to 14 years of age, called *General Basic Education*.

The new system branched out into: (a) a noncompulsory, academically oriented secondary education plan of study, usually leading to university study, called "Bachillerato" (GCE), and (b) a second, practically oriented plan, where studies related to the demands of the labor market and the need for technical training, called Vocational Training. Formally both branches were supposed to form a single Secondary Education System.

Among the changes introduced by the General Education Law (*GEL*) were those concerning the assessment of educational performance. Before the introduction of the GEL, the most important activity in the Spanish evaluation system was concerned with the administrative work of control and promotion. Therefore, in the official documents of the time we can find continuous references to terms such as "exam" or "test", "mark" or "qualification". The activity of evaluating, judging, controlling, and directing the work of students did not go further than giving a final qualification, the passing of a section of schooling.

The GEL tried to change this situation with a completely new orientation to student assessment. Among all the changes proposed, I will emphasize the following:

M. Niss (ed.), Cases of Assessment in Mathematics Education, 9–20.

- The assessment of performance refers to the pupil's development as well as to institutional actions.
- The assessment of the pupil's performance should take into account the formation and instructional level of each course as well as an appreciation of all aspects of the pupil's development and his or her aptitude for further study.
- The assessment should be carried out using a system of continuous assessment.
- A confidential record should be kept of the data and observations on the student's progress both of continuous assessment and of the set of formal tests, as well as of any other information necessary for the pupil's adequate orientation and education.
- The final mark in each course should include a qualitative appraisal, which may be positive or negative, and a weighted evaluation if the mark is positive.
- Quantitative marks are forbidden, and qualitative assessment is stated using a system of categories.
- Continuous promotion within the compulsory period of schooling is established; provision is made for pupils to spend a further year in the same class if they have not shown a sufficient command of the material.
- The assessment of pupils in Bachillerato courses will involve a joint marking system, carried out by the teachers in a collegial manner, following the criteria of the program in question and by assessment criteria laid down by the didactic seminars.

The Spanish school system adapted slowly, and at great expense to the assessment methodology introduced by the General Education Law of 1970, especially in three areas:

- The disappearance of quantitative marks, and their replacement by a system of categories.
- Continuous promotion within the compulsory period of schooling.
- The development of new methods and instruments to carry out a finer continuous assessment more closely related to the pupil's learning.

In the seventies the assessment model that was used was criterion-referenced, based on a more or less exact determination of the objectives to be achieved by the pupils, the elaboration of item banks appropriate to these objectives, and the administration of tests that measured whether the pupil had achieved the objectives set out.

In Spain the development of the General Education Law led to the establishment of a mathematics curriculum based on the "new mathematics". We can state that the model of operative objectives, together with an

emphasis on the formal and structural aspects of the organization and development of the contents, led to a clearly behavioristic conception of the learning of mathematics. This also affected the design and development of the assessment techniques and instruments used.

The emphasis was on the knowledge of facts and definitions and the performance of operative skills, among which we could emphasize the control of the properties used in each step, and the insistence on a formal explanation of each component used in the deductive reasoning process. The neglect of the use and practical application of mathematical knowledge reached alarming proportions. Spatial and plane geometry were totally abandoned. By the mid-seventies, this model was exhausted; because of the Spanish political situation, with its process of political reform and the establishment of democracy, efforts to improve education were limited.

At the beginning of the 1980s a review of the educational system was considered. The socialist government, which came to power after the 1982 elections, at first limited itself to a consideration of the renovation of examination papers and teaching programs. It has now embarked upon a much more ambitious project. Integration into the European Community has brought with it the need to extend the period of compulsory education to 16 years of age, an enormous change in the education system in force till now. A period known as "the Reform" began in 1986; the relevant characteristics of our future education system are now being subjected to a social and technical debate.

2. THE PRESENT SITUATION

In 1991, with the announcement of the new Law for the Organization of the Educational System, the bases and principles on which education in Spain is going to based in the years to come have now been established. Education is now compulsory from 6 to 16 years, and there are two different levels: Primary Education from 6 to 11 years, and Secondary Education from 12 to 16 years. Preschooling is open to all children from 3 to 5 years, and, in time, it will cover the years from birth to 5. The form this new system has taken was preceded by an extensive debate and consultation with the various social institutions. It has been made explicit in a series of programmatic documents published by the Ministry of Education and by the technical advisers to the Autonomous Communities. These documents are known as the *Basic Curricular Designs*.

The Basic Curricular Design (Ministerio de Educación y Ciencia, 1989) prepared by the national administration presents key ideas that nurture the Reform, including those concerning the new principles of assessment and evaluation. This document consists of four sections. In the first section, the theoretical and conceptual principles that have inspired the Reform are established. As a starting point, the existing educational system is described

and the need for a new frame of reference and the general characteristics of that frame of reference are described. The notion of curriculum, its functions, and the questions that should be answered are discussed; among these are the questions of "what, when, and how to assess". This section also establishes the distinction between curriculum design and curriculum development.

The second section is devoted to presenting the Basic Curricular Design, and indicating the functions that it fulfills, and the levels of responsibility and application that it expects. The constructive ideas on which the design is based are explained in detail; they emphasize the need to start from the pupil's level of development, the need to ensure the construction of significant learning with learner autonomy, and ways to modify the pupil's knowledge schemata using intense activity.

These assumptions have the following implications for assessment:

- Assessment enables us to collect information, to make value judgements, to orientate the teaching/learning process, and to make decisions about this process.
- The objective of assessment is to evaluate abilities.
- Abilities are expressed in the list of general objectives by school level and subject area.
- Abilities are not assessed directly, but indirectly, using the appropriate indicators.
- It is not the aim of assessment to measure behavior or performance
- Assessment should be continuous and individualized, it should be of a formative nature and should aim to establish valid criteria.
- The aim of assessment should be to orientate the pupil and to guide the teaching and learning process.
- The design also contemplates carrying out final assessment at the end of school levels or cycles.

Some further considerations are in order here: The previous sections are mainly of a theoretical character, there is no indication whatsoever of the way they should be carried out in practice. The document offers no method or tools for the "indirect" assessment of pupil's abilities. The lack of a concrete proposal for assessment makes the change very difficult, and the majority of classroom teachers still have to rely on the traditional method of "behavior or performance" to assess their pupils. As you go further up the levels of the school system, the reality of assessment is very different from what the design intends. Because of the social pressure to get an outstanding qualification, pupils give more importance to those aspects of education which may help them pass their standard tests successfully than those which may help develop their understanding and knowledge of a subject.

With respect to mathematics, it is clear that activities that lead to knowledge of facts, definitions, and concepts as well as the skills of calculation, reasoning, and representation, are predominant and make up practically all the activities of assessment. This is especially pronounced in the final courses that end the period of secondary education, where the tradition of equalizing assessment with exams, evaluation with qualification, and orientation with promotion are strongly established. It is difficult to introduce other possibilities.

In the third section, the components of the Basic Curricular Design are described at the levels of stage and subjects area. In the curricular design each subject area which makes up the curriculum of compulsory education, including the area of mathematics, should consider four fundamental components:

- Objectives
- Contents
- Methodology
- Assessment

These four components make up a system presenting interrelationships which must be emphasized and developed; they cannot be considered in isolation. The objectives establish the abilities which should have been acquired after the initial educational periods. They refer to five types of human abilities: cognitive, motor, affective, interpersonal, and social. The contents include: concepts, facts and principles, procedures, values, norms and attitudes.

In this third section, the field of competence assigned to the school is also established and should take shape in a document called *Project and Curricular Programming* The document explains the set of decisions that are made with respect to what, how, and when to teach and assess. This seems to disseminate the ideas and provides coherence and identity to regional centers.

Finally, a fourth section is dedicated to indicating lines of action that are of high priority to the administration. Six lines are presented: teacher training, curricular materials, support services for the school, organization of centers, education research, and assessment. With respect to assessment, three levels are established: pupils, centers, and the educational system. The principle of continuous promotion of pupils is specifically recognized, as well as the conditions for moving into the next cycle and the certificates that will be obtained in each case.

The assessment of the centers is entrusted to an inspectorate service with specific functions. The service recognizes that the curricular project of the center conducts its assessment, but is concerned with the assessment of the education system. Its document, valid for the whole compulsory education period, follows the general principles that mark the path Spanish education

is to follow. It includes a more particularized section where each of the stages, primary and secondary, of the system are presented. These sections describe the general characteristics of each stage, its objectives, its curricular structure, and its corresponding didactic guidelines. The document includes a presentation of each of the areas of knowledge that make up the curriculum of each stage. In each area there is an introduction that presents the principles around which its development will be articulated. The specific objectives of the area, the blocks of content, and didactic and assessment guidelines are stated.

3. THEORETICAL FRAMEWORK OF ASSESSMENT IN MATHEMATICS

The fundamental ideas that appear in the guidelines for assessment in the area of mathematics (Rico, 1990) are the following:

1. Reasons for assessment

- We have to carry out systematic observations so that the teacher can make judgements about the progress of the learning process.
- Assessment is an integral and fundamental part of the teaching and learning process. Assessing performance enables the teacher to control and improve it. The pupils' reflection on their achievement problems help them control and get involved in the learning process.
- Assessment has to consider attitudes and general procedures. To do this we must modify the usual techniques and instruments.
- Assessment is not a goal in itself; it must be continuous and differentiated for each individual pupil.

2. Self-assessment of pupils and teachers

- Self-assessment requires pupils to carry out a critical reflection on the learning process and take responsibility for their education; self-respect and independence are also fundamental.
- The observation, assessment, and adjustment of the teacher's performance is a key factor in the teaching/learning process.

3. Instruments for observation and assessment

- The recording procedure should be simple and not require much time. One card per pupil should be used to note observations about how attainment of the learning objectives is shown; the results of specific tests should also appear on the card.
- The observation of each pupil should be carried out on a regular basis, and criteria established to guarantee this. Discussions in class give an opportunity to appreciate the pupil's ability to argue coherently, their command of vocabulary, and the respect they show to others.

- The class notebook can be another source of information; the activities carried out by the pupil should appear in it: exercises and problems, summaries and diagrams, etc. The data which the notebook provides are the level of critical and graphical expression, work habits, etc.
- We can also get information by carrying out specific assessment activities. There are a wide variety of types of tests with advantages and disadvantages. It is advisable to select those that provide a multitude of possibilities to draw out the initiative and ability of the pupils.
- To fulfill the orientative aim of assessment we must inform the pupils of the successive evaluations that have been carried out on their learning process, indicating to them alternatives for remedial work — if necessary — and pointing out achievements and progress.

This theoretical framework does not describe the real situation of assessment in Spain at the present time. Other considerations will enable us to grasp more accurately the status of assessment in mathematics in the school system.

4. CONDITIONS THAT AFFECT ASSESSMENT IN MATHEMATICS

In the current situation, assessment in mathematics in Spain has the following features:

1. The most innovative approaches, that are connected to the most advanced currents within mathematics education, have been put forward in the framework of renewal and change to reform the education system. Some of the important ideas that have contributed to the curricular design for mathematics are the developmental framework for constructing mathematical knowledge, the importance of inductive reasoning and intuitive procedures in the work of mathematicians, the potential of mathematics as an instrument of communication, and the essential constructive aspect of the elaboration and acquisition of mathematical knowledge.

Learning mathematics is considered to encourage the development and acquisition of very general cognitive abilities, but emphasis should also be put on utilitarian and pragmatic aims such as mathematical needs in adult life. In this same way participation of girls is especially fostered. Cognitive psychology has been considered in the way in which the contents and the different competencies that are indicated in the objectives are classified and organized.

We can say that the mathematics curriculum design has been inspired by the most well-known and respected trends in the communities of Anglo

Saxon mathematics educators (Romberg, 1989). There is a strong tendency to give the cognitive competencies derived from the procedures and strategies necessary for problem solving greater value than they had been given previously. The traditional curriculum of Spanish mathematics, heavily influenced by the French and Central European rationalists and structuralists, has been considerably modified by empirical, pragmatic, and procedural ideas coming from the Anglo Saxons and, partially, from the Dutch.

2. We should take into account that the majority of the teachers working today were trained during the 1970s and so their training was structuralistic, with a great emphasis on formalism, on correction of procedures and on conceptual control using definitions and symbolic notation.

Before the 1970s, the majority of secondary teachers in charge of this subject were not specialized in mathematics. This situation fortunately changed lately. Nowadays most secondary teachers — more than 80 percent — have a degree in mathematics (university graduates with 5 years of training). Yet, even though the technical knowledge of secondary mathematics teachers has improved, the psycho-pedagogic training is nonexistent or very deficient. Their only source of information is their own teaching experience. Only a few small groups have worked in a systematic way on didactic problems, taking psycho-pedagogic training to some very worthy levels, but having a limited influence.

However in the past few years, a scarcity of university graduates in mathematics who could occupy secondary education teaching posts has been noted; it has begun to be normal for a graduate in chemistry or biology to teach mathematics at these levels. On the other hand, primary teachers have training in mathematics that is sufficient to teach at the primary level. The knowledge they have is different in nature from the knowledge that a university graduate has, and they have very little awareness of the specific nature and use of mathematics in their own field of work. Even when the psycho-pedagogic training that primary teachers have received is considerable, the connection between this training and the role that they should play as mathematics teachers is not well established So we find ourselves in a situation where the present members of the teaching profession come from two very different types of training, each having one very well-developed component and the other very weak or nonexistent. Both types are, to a certain extent, complementary; however, a systematic collaboration between them which would have been profitable for both parties has not been favored.

Although the approaches of the reform are advanced and are connected to the most up-to-date current in mathematics education, present teachers are not prepared collectively to take over this task. The prospects for the future are even more dismal since the decrease of the numbers of university graduates in mathematics who are preparing to become

secondary teachers is accelerating. On the other hand, the mathematics training of future primary teachers is not going to improve immediately. The present administration does not have a specific plan for the basic preparation of mathematics teachers. Because of this and the opportunities offered in the labor market, mathematics teaching post are in the future going to be covered by graduates who have neither adequate preparation as teachers nor as mathematicians.

3. The current model of assessment is centered basically on paper-and-pencil tests in which pupils are to show their command of the facts, skills, and definitions that make up the most fundamental and simple aspects of mathematics knowledge.

It is very rare that the pupils are presented with creative activities or that their competence is assessed when they confront tasks that they have not previously tried and in which they have to put to use all their knowledge of one specific subject. Even when groups and teams have worked on innovation in mathematics teaching/learning, the majority of their efforts have been centered on determining and clarifying the objectives, organizing the contents or incorporating new topics into the corresponding levels of teaching and above all, elaborating a wide variety of treatments, adaptions, resources and planning of activities. Such work may be used to determine the treatment and methodology with which to control the content to facilitate learning.

Systematic work on innovation in assessment is very scarce, and its diffusion to groups other than those who developed it, is even more scarce. The innovations in the other components of the mathematics curriculum (objectives, contents, and methodology) have not influenced the development of new approaches in assessment.

4. The term *evaluación* has been contaminated strongly in the Spanish educational system, it is usually identified with exam, final test, or mark. As the pupil goes through the system, the weight of assessment lies increasingly on the final examination as an administrative act. Pupils of any level immediately identify assessment with exam, with promotion, and with control.

A survey carried out recently with pupils from different levels of the education system showed the identity between assessment and a final test, and the identity was strongly criticized for its limited and deficient character. Pupils perceive the institutional system of assessment as having no other aim than that of supervision, which they reject. Even in that minority of pupils who feel gratified by their good results, there is no doubt that the main aim of the assessment is the control of the pupils' knowledge.

5. At the congress on Mathematics Teaching and Learning, which took place in Castellón in March, 1991, the present system of assessment of

mathematics in secondary teaching was characterized by the Working Group on Assessment in the following way:

- There is a rigid pattern of timing, since the assessment is centered on one or two written tests each term, with some weeks dedicated exclusively to carrying out examinations or reexaminations.
- The explicit aim of the tests is to give a course mark.
- The overall character of the marks given to the pupils is that of a summary of different aspects and information obtained with different exercises; the complexity of the learning achieved by the pupils is masked by assessment that yields one item of information.
- The level of an acceptable command of the knowledge is indicated by an arbitrary line, which is called the "pass level" or "to have a five" (i.e., to get 5 out of 10).
- Neither the pupils' mistakes nor their unanswered questions are in any sense evaluated.
- Competency tests are set in parts, and in most of the cases the contents of the parts that are passed are not examined again.
- There is a compulsory retest in September for a considerable number of unsuccessful pupils, those who do not pass the retest must repeat the course.

At this same congress, the teachers in the Working Group, after a long debate about the criteria to be used for assessment, pointed to the following ideas as goals to be aimed for:

- Assessment should consider the pupil's ability, not only at the end of the course but at the beginning.
- Assessment should be a continuous process, providing reliable information about the progress and deficiencies of the pupils; it should serve as feedback to pupils and teachers.
- Assessment should be a constant activity in the teaching/learning process; we must abandon the ritual of isolated and particular acts linked to assessment.
- The formative character of the assessment process must be underlined, and punitive connotations taken away; assessment cannot be equated with the final mark.
- Assessment should affect not only the pupil but also the teacher and those other elements of the education system that contribute to the pupil's achievement.
- The assessment of each pupil should be carried out on a personal and individualized level, taking into account that it is not always possible to evaluate all the pupil's learning. On the other hand, we must not only evaluate the pupil's command of concepts but also his or her attitudes and procedures.

5. CONCLUSION

The intellectual restlessness of the Spanish teachers in General Basic Education with respect to assessment is an indicator of their awareness that at the moment there is a strong need to orientate the evaluations and judgements of the teacher in a direction that will contribute to effective learning and to the development of pupil's self-esteem, communication ability, and social integration. We can perceive that it is a vitally important field of work, and in this sense, the community of mathematics education in Spain is beginning to work systematically and in a coordinated fashion, contributing to the diffusion of individual and group initiatives in assessment. This has been shown in the Working Group on Assessment in Mathematics that met in Castellón (1991), and in the studies that are in progress in Barcelona, Valencia, Salamanca, Madrid, Zaragoza, Tenerife, Sevilla, Málaga, Granada, and many other towns throughout Spain.

Although it is a very difficult task, there are key ideas that should orientate the future work. Of these ideas we want to underline the following ones:

First, we should consider assessment as a continuous and interdependent process with the other components in the curriculum; contents, objectives and methodology cannot be dealt with as isolated questions in the process of assessment, but rather must be contemplated as interconnected. Assessment is not an isolated single element but one that should impregnate all the stages that make up mathematics teaching and learning.

Second, the formative and orientative character of assessment is another idea that has to be developed; assessment should be considered a critical judgement that stimulates, orientates, and promotes a better understanding and a greater control of knowledge on the part of the pupils, that shows them the mistakes and defects of their work, that marks their path to success, and that makes them feel satisfied with the effort. The teacher should stimulate and develop this style of working on a day-to-day basis.

And finally, we need to use a variety of methods and instruments, some which are systematic, and others that favor the creative aspects of mathematics, to show the many facets of a rational organization of knowledge, and to stimulate invention. This result of the effort and evaluation of the teachers will serve not only to encourage the development of the pupil's potential, but also to document it.

NOTE

The present study was commissioned by the Spanish Federation of Mathematics Teachers' Societies, following a proposal made by the Andalusian Society for Mathematics Education,

"Thales". My thanks to Neil Maclaren, Luis Quereda and Jeremy Kilpatrick for their assistance with the English translation.

REFERENCES

Ministerio de Educacion y Ciencia: 1972, *Ley General de Educación*, Servicio de Publicaciones del Ministerio de Educación y Ciencia y Boletin Oficial del Estado, Madrid.

Ministerio de Educación y Ciencia 1989, *Diseño Curricular Base. Educación Primaria y Educación Secundaria Obligatoria*, Servicio de Publicaciones del Ministerio de Educación y Ciencia y Boletin Oficial del Estado, Madrid.

Rico, L.: 1990, 'Diseño Curricular en Educación Matematica Elementos y Evaluación', in Linares, S. & Sanchez, V (Eds.), *Teoria y Practica en Educación Matemática*, Editorial Alfar, Sevilla.

Romberg, T.A., 1989, 'Evaluation A coat of many colors' in Robitaille, D (Ed.), *Evaluation and Assessment in Mathematics Education*, Science and Technology Education Series No. 32, UNESCO, Paris.

Luis Rico
Departamento Didactica de la Matemática,
Universidad de Granada,
Spain

DESMOND R. BROOMES & JAMES A. HALLIDAY

MAJOR ISSUES IN ASSESSING MATHEMATICS PERFORMANCE AT 16+ LEVEL: A CARIBBEAN PERSPECTIVE

1. INTRODUCTION

From London, England to Bridgetown, Barbados

The Caribbean Examinations Council (CXC) under a mandate from fifteen territories of the English speaking Caribbean region (Anguilla, Antigua & Barbuda, Barbados, Belize, British Virgin Islands, Dominica, Grenada, Guyana, Jamaica, Montserrat, St Kitts-Nevis, St Lucia, St Vincent, Trinidad & Tobago, and the Turks & Caicos Islands), has been constructing and administering secondary school certificate examinations since 1979. These territories are members of the regional organization, the Caribbean Community for Economic and Social Development (CARICOM). They spread across the Caribbean Sea, blue and green, in a crescent of islands and two mainland territories from Belize in the northwest to Guyana in the south, cover a total land area of about 258,000 km^2 and have a population of 5.5 million approximately.

These English speaking territories which were former colonies of the United Kingdom, modelled their education systems to a very large extent, on the education system of Britain. After five years of secondary schooling, students wrote external examinations such as *General Certificate of Education* (*GCE*) Ordinary Level of London University or Cambridge University. The more able students continued schooling for a further two years and wrote the Advanced Level examinations of the same universities. Passes in five subjects at the Ordinary Level and two at the Advanced Level were usually accepted by universities in the United Kingdom and in North America as matriculation requirements.

The GCE syllabuses and examinations as set by British institutions usually reflected the cultural biases and educational philosophies of Britain. Consequently, some of the course outlines and emphases were irrelevant to the needs of the Caribbean territories. It was natural, therefore, that as these territories moved towards political independence, the urge to develop educational systems and secondary school examinations more relevant to their needs and aspirations became critical. CXC was therefore established in 1972 to provide, through the activities of Caribbean teachers and

M. Niss (ed.), Cases of Assessment in Mathematics Education, 21–33.

Caribbean economic and cultural institutions, relevant secondary school syllabuses and examinations to replace those of the *British Overseas Examinations Boards*.

The impact of CXC's examinations on the curricula of secondary schools in the Caribbean region was profound. Schools expanded their curricula both quantitatively and qualitatively. In 1979, CXC first offered examinations in five subjects. By 1990 CXC offered thirty-three subjects. In 1979, 30,276 candidates wrote 61,584 subject examinations; in 1990, 73,540 candidates wrote 281,599 subject examinations. In particular, over the same period, candidate entries in mathematics grew from 19,805 to 50,034.

2. THE EXAMINATION-CURRICULUM DIALOGUE

School certificate examinations tend to exert tremendous influence on individual students, on schools, and on the community as a whole. Many of these effects are salutary and beneficial to the individual and society, provided that the quality of the tasks used to assess students and hence to evaluate school programs, meets certain established psychometric and ethical criteria.

CXC seeks to construct quality tests having certain desirable psychometric properties by employing, so to speak, a research design with a set of independent variables. The principal independent variable that is manipulated by CXC in designing tests for school certificate examinations, relates to the *nature* and *size* of the correspondence that exist between the domain of instructional activities actually delivered to students and the domain of content included in the tests. The domain of instructional activities is usually defined as the school program and partly described in a CXC syllabus.

CXC syllabuses were developed in a rational and systematic way. A major CXC policy was to develop syllabuses which were sensitive and responsive to specific regional needs, and to establish testing practices which were informed by the extant progressive ideas from the international measurement community.

The syllabus development in each subject area was set in motion through a series of meetings and consultations with practicing teachers, curriculum specialists, measurement experts and distinguished educators in the Caribbean region. Further technical advice came from renowned international educators and testing institutions. For example, in the formative years, Caribbean examiners performed examination tasks under guidance of London and Cambridge examiners, while technical advice and training in measurement and testing were obtained from Educational Testing Services, Princeton, USA, and the Ottawa Testing Unit, Canada.

In many ways, therefore, CXC syllabuses and examinations sought to exploit the existing know-how about syllabus construction and test

development and to go beyond by imposing a Caribbean mould on what was being done, how it was being done and why it was being done.

First, the syllabuses were developed under two schemes: A *General Proficiency* scheme and a *Basic Proficiency* scheme. This was a deliberate strategy to address the instructional and organizational limitations of secondary schools whose main objective hitherto was in terms of preparing candidates for white collar jobs and university entrance requirements.

Second, the syllabuses were written to guide and inform the instructional program within each school. For example, the mathematics syllabus contains not only a list of topics arranged as sequence of subtopics, but also objectives stated behaviorally and expressively, and tasks illustrating and defining the required content and processes. CXC syllabuses are intended to establish communication links among teachers, curriculum developers and test constructors.

Third, and as a corollary to the above, CXC syllabuses inform the teacher, the candidate and the Chief Examiner (qua test constructor) about the rules under which the examination papers are set, marked and graded; that is, the number of items on each topic, weighting of papers, weighting of profile dimensions etc. of the examination.

The second and the third sets of features described above are critical for criterion-referenced testing. They help to prescribe a domain sufficiently well defined, that teacher, test constructor and even the candidate can, with some degree of congruence, identify the tasks of the examination. These features control the test design and contribute to validity, reliability and generalizability of the test scores.

Controlling and manipulating this test design variable has generated innovative responses to the problems of assessing students' performance in mathematics. These problems have many ramifications.

First, a CXC syllabus sets down the *content* and *processes* in terms of specific *objectives*. It also sets down content and processes as expressive objectives. However, it is not always easy to identify what to test and how to test whatever is being tested. The technology of objectives (Bloom, 1956, Bloom, Hastings & Madaus, 1971) has enhanced the work of the test development community very much. Nevertheless, our research is pointing up certain limitations in the technology. Simply put, test constructions seems to favor the operational definition of an objective which is linked not to a network of meaning but to an automatic conditioning. Carefully and gradually, new research questions that impinge on the meaning of knowledge are emerging.

Second, the difficulty of setting questions to match compound and complex objectives such as the following (taken from CXC Mathematics syllabus) is a real one:

- Use certain number properties and concepts in simplifying computational tasks (Number Theory, Specific Objective 10, p. 14)

- Use linear equations and inequalities to solve word problems (Algebra, Specific Objective 12, p. 17)
- Use matrices to solve simple problems in geometry (Vector Matrices, Specific Objectives 10, p. 26).

Considerable mismatch between test items and objectives is likely to result. Griffith (1985) using an index of item-objective congruence as devised by Rovinelli and Hambleton (1977), found that content specialists in mathematics and CXC test constructors were able to assign only 56 per cent of a 50-item mathematics test to the same dimensions.

Third, a test designed to measure higher order thinking ought to include items which contain situations unfamiliar to the candidate; make use of novel conditions, and; employ test formats in ways different from those used during instruction.

Herein lies a dilemma. To be fair to students, the test items need to be close to those used in the instructional program. To measure the higher order learning objectives with fidelity, the test need to be far removed from those of the instructional program. Linn (1983) identifies this paradox as a key problem that must be addressed by those responsible for testing and instruction. He puts forward four considerations that may create useful links (content match, use of feedback, a flagging function, and attachment of sanctions and rewards to results). However, they are limited to single classrooms or at best, the individual school.

Fourth, students writing CXC examinations in mathematics are, so to speak, nested within schools and schools are nested within territories. Thus a given student's opportunity to learn how to perform any relevant mathematics task, depends heavily on factors outside his/her control. In the Caribbean context, these factors may include lack of specific resources in a school (geometrical models, calculators, library materials), omission of the teaching to deliver appropriate instruction, and inability of instructional program to address the mathematical processes of the syllabus (size of class, ways of organizing classes, time available, teaching skills). It is manifestly unfair to the student that he/she should be required to show competence on mathematical topics and processes he/she has had no opportunity to acquire.

A way out of (i) having no opportunity to learn the desired abilities and attitudes, and (ii) the validity dilemma of what is on the test and what has been taught, suggests that each student should be personally responsible for what should have been taught and how it should have been taught as set down on the CXC mathematics syllabus. This implies a tremendous, albeit a necessary burden on an Examinations Board to produce the operational definitions for the topics included in the examination syllabus. In a sense, examination boards need to engage in curriculum development and maintenance at the classroom level.

Airasian & Madaus (1983) in an article that considers policy and practice in terms of the appropriate interface between tests and instruction in a North American context, conclude that

"if the test is used for individual certification, then fundamental fairness seems to dictate that it is not enough to show that on average, throughout the state or district, adequate opportunity to learn was provided. It would seem to be necessary to show that each pupil received timely and adequate opportunity to learn." (p. 114).

3. STRATEGIES USED BY CXC TO CREATE LINKAGES BETWEEN CURRICULUM AND TESTING

CXC addresses the problem from at last three fronts using the experience and expertise of the teacher aggregated across schools and across countries.

First front: teachers (from all 15 contributing countries) help to define the content and processes of the CXC Mathematic syllabus. They make suggestions on the structure and content of the syllabus on an ongoing basis. Further, each year, they comment on the match between the syllabus and the test questions, and on the goodness of each question. At syllabus review, they give evaluative comments on the examination questions as valid measures of the syllabus objectives.

Second Front: teachers participate in defining and constructing test items used to measure the learning objectives of the syllabus. At test construction workshops, national and regional, teachers prepare examination questions and review them. Over the years a cadre of teachers trained in test construction techniques has emerged. A major advantage of this strategy is that classroom teachers who study how children at 16+ think and learn, bring to bear their experience and knowledge to inform the nature of the tasks to be included in the tests that CXC should set. One disadvantage, however, is that teachers sometimes tend to have inflated expectations of what children can do under examination conditions.

Third Front: teachers review the mark scheme for the examination papers, mark examination scripts and participate, in a limited way, in the grading of scripts. Their involvement has proved invaluable and on several occasions their insights have led to major modification of mark schemes, thereby enabling chief examiners to assess candidates validly. The exposure of teachers to these activities also helps them to revise their curriculum practices and improve the quality of their teaching. Thus, a deliberate thrust of CXC's activities is to integrate testing and instruction and to use tests as instruments for instructional improvement.

In the early years (1979–1989) CXC emulated, and for cogent reasons, the practices of conventional test development: first, tests are developed to assess a construct or a trait (stated as an objective); second, test items are selected from a bank of items or are constructed; third, data on the

performance of the items are collected under special conditions including opportunity to learn the desired content; finally, items are accepted, revised, or rejected on the basis of the goodness of fit between the item data and the hypothesized model of mathematical performance. In the words of Baker & Herman (1983), the psychometrician

"remains outside the action of instruction. His focus is on fidelity rather than improvement. The measurer should make no ripples." (p. 150).

Gradually, CXC is departing in major ways from this conventional wisdom. The press for improved mathematics teaching and learning in all Caribbean schools has become dominant. In the mind of the Caribbean educator, the measurer must make ripples. That is, CXC and its testing activities must be intimately integrated into curriculum development and teacher education. This developing thrust of the 1990's requires at least two major activities. First, CXC must specify and delineate an *inside* view of learning mathematics and also of teaching mathematics. Secondly, CXC must conceptualize the role of measurement in terms of the inside view and define the nature of the test content and processes from the inside view.

CXC's search for ways of accessing the performance of a larger proportion of the age cohort forced a realization that the psychological constructs used to describe and explain the student's mathematical knowledge might usefully be described in the terms of three phases.

Phase 1 focuses on the need for the individual to possess a critical mass of information and so requires the individual to encode and store information, access information he has previously stored, interpret data in terms of the individual's present knowledge and beliefs as well as to draw out relationships in order to strengthen the present knowledge structure and increase its capacity to accommodate more data.

Phase 2 stresses the ability and willingness of the individual to structure newly acquired information, link conceptual knowledge to certain skills and refine the skills, as well as to identify and structure general cognitive abilities.

Phase 3 is characterized by a variety of representational skills, restructuring skills, and timing skills and most importantly by flexible applications of a variety of specialized schemes to problem representation and solution.

4. NATURE OF ACHIEVEMENT

The research evidence during the 70's and 80's (Anderson, 1982, Chiesi, Spilich & Voss, 1979, Chase, 1973) supplied the data which enabled cognitive theorists to identify at least three distinct phases of cognitive learning and to describe the prominent processes of each phase. The above specifications have made use of the research data. "Thus achievement is as

much an organizational function as it is an acquisition function" (Snow, 1980, p. 42). Hence, educational achievement, and by definition educational measurement, should focus not only on storage and retrieval of information, and interpreting and processing of information, but also on the nature of the scheme used in interpreting and processing information, the reorganization of schemes into patterns, networks, hierarchies, etc. and the application of variety of perspectives to problem representation and solution; and further, on the initiation, regulation and monitoring of action, and finally on the reflection on and evaluation of action and thought.

The individual's knowledge structure therefore may usefully be described to embrace:

- *Conceptual knowledge* (procedures for recognizing patterns)
- *Procedural knowledge* (skills the individual can do, specifications he can carry out)
- *Strategic knowledge* (setting goals and subgoals, forming plans for attaining goals)
- *Problem solving.*

Thus CXC, in reporting on a candidate's mathematical performance, reports mathematics achievement under three profile dimensions: *Computation, Comprehension, Reasoning.* These three profile dimensions were derived from the three levels of mathematical thinking, Recall, Algorithmic Thinking, and Open Search as defined by Avital and Shettleworth (1968) and motivated by the work of Bloom (1956).

Candidates' performance in the CXC Mathematics examinations is described in two main ways:

- as an *overall grade*, where
  Grade I denotes a *comprehensive* knowledge of *all* aspects of the syllabus
  Grade II denotes a *working* knowledge of *most* aspects of the syllabus
  Grade III denotes a *working* knowledge of *some* aspects of the syllabus
  Grade IV denotes a *limited* knowledge of *few* aspects of the syllabus
  Grade V denotes that the candidate has not produced sufficient evidence on which to base a judgement

- as a *profile grade* where Grade A denotes "above average"; Grade B, "average"; Grade C, "below average"; and NA, "no assessment possible".

Thus candidates' performance in CXC Mathematics examination may be reported as follows:

Grade I: Computation (A), Comprehension (A), Reasoning (A)
Grade I: Computation (B), Comprehension (A), Reasoning (A)
Grade II: Computation (B), Comprehension (A), Reasoning (B)
Grade II: Computation (B), Comprehension (B), Reasoning (B)
Grade III: Computation (C), Comprehension (B), Reasoning (C)

The *raison d'être* for reporting candidates' performance using an overall grade and a profile-dimension grade derives from an appreciation of the link between instructional and testing practices and also between occupational requirements and certifications. Halliday (1989) listed six major uses for profile reporting as introduced by CXC. Five of them related to the match between curriculum and testing.

5. CXC SPECIMEN ITEMS AND QUESTIONS

Sample Items from the Multiple-Choice Papers

The following multiple choice items illustrate some of the major points about CXC testing procedures which have been highlighted in this paper.

The score which occurs most often in a distribution of scores is the
(a) mean
(b) median
(c) mode
(d) frequency

Figure 1 *Item 1*

What percent of 20 is 16?
(a) 21%
(b) 36%
(c) 75%
(d) 80%

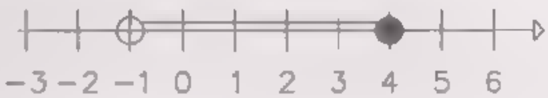
Figure 2 *Item 2*

Item 1 requires the candidate to recognize a concept he/she previously learned. Item 2 requires the candidate to do a simple computation of the type that he has done repeatedly in class. It requires no more than a one-step calculation. These two items are classified as measures of the profile dimension, *Computation*. Items of this level of difficulty are used to define the grad IV/V boundary.

Item 3 tests conceptual knowledge and requires the candidate to translate from one mathematical mode (diagrammatic) to another mode (symbolic). It is classified as a measure of the profile dimension, *Comprehension*. An item of this level of difficulty would define the Grade III/IV boundary.

Item 4 requires conceptual knowledge of rectangles, procedural knowledge of calculating the area of rectangles, and strategic knowledge of how to formulate a problem, search for a solution and test the solution. It is classified as a measure of the profile dimension, *Reasoning*. An item constructed under these specifications would define the Grade II/III

boundary. This is the boundary usually used as selection criterion for entry into university.



The graph of the inequality in the diagram above is defined by

(a) $-1 \leq x < 4$
(b) $-1 \leq x \leq 4$
(c) $-1 < x < 4$
(d) $-1 < x \leq 4$

Figure 3 *Item 3*

The area of a rectangle is 14.4 cm². If the length is multiplied by four and the width is halved, the area would then be

(a) 7.2 cm^2
(b) 14.4 cm^2
(c) 28.8 cm^2
(d) 57.6 cm^2

Figure 4 *Item 4*

A bag contains red and blue marbles of the same size and mass. There are 8 blue marbles. If the probability of drawing a blue marble at random is 2/5, the number of red marbles in the bag is

(a) 10
(b) 12
(c) 20
(d) 32

Figure 5 *Item 5*

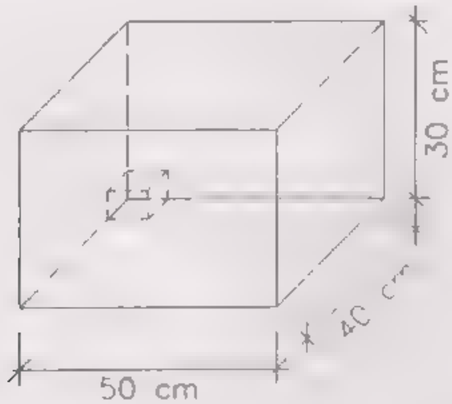
Item 5 is another example of a measure of the profile dimension, *Reasoning*. The item serves to define the I/II boundary. A study of the cognitive demands of this item shows that it requires more complex abilities for its successful completion than Item 4.

Sample Questions from the Essay Papers

Question 1 requires the candidate to decode the information in the diagram, access relevant information previously learned about the volume of cuboids, and the equation between liters and cm^3 and draw out the relationship between volume and base area in order to solve the problem.

To do Question 2 successfully, a candidate is required to represent a real word problem in a form that is amenable to mathematical treatment, to show the relation between a set and a subset as a fraction, a ratio and a percentage, to do mathematical calculations and interpret the results of these calculations in the context of the problem. Thus a candidate must demonstrate competence in a variety of skills in order to solve the problem successfully.

Question 1: (CXC Mathematics Basic Paper, 2. June 1989)
Question 5 part (a)



(a) The figure above, not drawn to scale, represents a fish-tank in the shape of a cuboid of height 30 cm.
(i) Calculate, in cm^3, the volume of the tank.
(ii) If there are 40 liters of water in the tank, calculate the height of the water in the tank.
(5 marks)

Figure 6 *Sample question from the Essay Papers*

Question 2: CXC Mathematics General Paper 2. June 1991
Question 1 part (c)

(c) The sum of \$2,500 is divided among Peter, Queen and Raymond. Raymond received half, Peter received \$312.50 and Queen received the remainder
Calculate
(i) Raymond's share
(ii) Queen's share
(iii) the ratio in which the \$2,500 was divided among the three persons
(iv) the percentage of the total that Peter received.
(5 marks)

Figure 7 *Sample question from the Essay Papers*

The CXC strategy of setting examination questions enables the examiners to construct a marking scheme that would allow marks to reflect the abilities described by the profile dimensions. To successfully do test items as given above, candidates must have a variety of skills and abilities. In some cases the abilities and skills needed are essentially abilities to do calculation or recall specific learned material — Profile 1, Computation; abilities to do translation — Profile 2, Comprehension; and abilities to solve problems — Profile 3, Reasoning.

Briefly, this strategy enables the examiners to define performance not simply in terms of marks, but in terms of abilities displayed. This strategy

therefore requires close linkages between what is done in schools and what is tested in the CXC examinations.

6. SUMMARY AND CONCLUSION

Assessing mathematics performance of secondary school students in the Caribbean provided major challenges to the Caribbean Examinations Council, national curriculum development units, teacher training institutions, and schools (primary and secondary). And the way these challenges have been accepted and are being resolved seems to be derived from

- an awareness of *validity* as the most critical feature of examination grades in mathematics,
- an understanding of validity as the extent to which different logics and languages of mathematical competence, confirmed by empirical evidence from various sources, can be used to support the *appropriateness* and *adequacy* of the influence made from the test scores, and
- a commitment to *linking* instructional practices to assessment procedures and to using a cybernetic mechanism.

Thus the paper showed that secondary teachers played pivotal roles in developing syllabuses, in constructing tests and in grading students' responses, that teachers, as they engaged in certain activities directly associated with the CXC examinations, upgraded their curriculum practices and improved their instruction in mathematics; that CXC test constructors and chief examiners acquired new insights about test construction through their interaction with curriculum and teacher training.

Further, the validity of the test scores as well as their reliability and their generalizability were substantially enhanced through

(a) *profile reporting* of the candidates mathematics performance in terms of three dimensions;
(b) constructing the mathematics *examinations* under two distinct schemes, Basic Proficiency and General Proficiency;
(c) constructing test items to fit specifications that make for the emergence of *mathematics competence*; and
(d) using three distinct modes of measuring mathematics competence — *multiple choice*, *short essays* (problems) and *extended essays* (problems).

A way forward for CXC should encompass:

1. *Refining* the technology of the *test construction* so that tests reflect the way mathematics achievement is being conceptualized (by

teachers and curriculum specialists) along a novice-to-expert continuum.

2. *Identifying* and defining *psychometric procedures* appropriate to criterion-referenced measures when mathematical achievement is being assessed across schools and across territories.
3. *Refining* the operational definitions of the *profile dimensions* used by CXC in mathematics so that:
   - test constructors would be able to design and classify mathematics test items/questions with more accuracy and fidelity,
   - students' performance could be described in ways consonant with their mathematical behaviors in class and in examinations.
4. Using a *wider* variety of *assessment procedures* (including school based assessment) to measure mathematics achievement, in keeping with the practice and philosophy of CXC.

REFERENCES

Airasian, P.W. & Madaus, G.F.: 1983, 'Linking testing and instruction: Policy issues', *Journal of Educational Measurement*, **20**, 103–118.

Anderson, J.R.: 1982, 'Acquisition of cognitive skills', *Psychological Review*, **89**, 369–406.

Avital, S. & Shettleworth S.: 1968, *Objectives for Mathematical Learning: Some ideas for the teacher*, OISE Bulletin No. 3.

Baker, E.L. & Herman, J.L.: 1983, 'Task structure design: Beyond linkage', *Journal of Educational Measurement*, **20**, 149–164.

Berk, R.A. (Ed.): 1980, *Criterion referenced measurement: the state of the art*, Johns Hopkins University Press, Baltimore.

Bloom, Benjamin (Ed.): 1956, *Taxonomy of Educational Objectives — cognitive domain classification of educational goals* Handbook 1, Longman, New York.

Broomes, D.R., Shepherd, E. & Dennison J.: 1986, *Defining and Constructing Questions for CXC Basic Proficiency Examinations in Mathematics*, University of the West Indies, Faculty of Education, Cave Hill Campus, Barbados.

Caribbean Examinations Council: 1988, *Secondary Education Certificate Mathematics Syllabus*, CXC, Barbados, WI.

Chase, W.G. (Ed.): 1973, *Visual information processing*, Academic Press, New York.

Chiesi, H.L., Spilich, G.J. & Voss, J.F.: 1979, 'Acquisition of domain-related information in relation to high and low domain knowledge', *Journal of Verbal Learning and Verbal Behavior*, **18**, 257–273.

Cronbach, L.J., Gleser, G.C., Nanda, H. & Rajaranam, H.: 1972, *The dependability of behavioral measurements*, Wiley, New York.

Dillon, R.F. & Stevenson Hicks, R.: 1983, 'Competence versus performance and recent approaches to cognitive assessments', *Psychology in the Schools*, **20**, 142–145.

Ernest, P.: 1984, 'The CXC Mathematics Examination: A Caribbean Innovation in Assessment', *Educational Studies in Mathematics*, **15**, 397–412.

Glaser, R.: 1963, 'Instructional technology and the measure of learning outcomes', *American Psychologist*, 18, p. 519–521.

Griffith, Stafford: 1985, *Convergent and Discriminant Validation of Measures of Subject Dimensions in Three Selected Examinations of the Caribbean Examinations Council*, unpublished doctoral dissertation, University of the West Indies, Jamaica, W.I.

Halliday, J.A. 1989, 'An exploratory study of the construct validity of CXC Basic Mathematics Examinations', MA Thesis, University of Toronto.

Halliday, J.A. 1990, *Measurement and Evaluation Report to Mathematics Panel on June 1990 Mathematics Examination, CXC*, unpublished.

Halliday, J.A. & Rose, G. St C. 1986, 'Error Analysis of Candidates' Responses to Selected Questions on CXC 1985 Basic Proficiency Mathematics Paper II', in Fraser-Abder, P. (Ed.), *Report on the first Biennial Conference on Science Education Research in Latin America and the Caribbean*, 201–218.

Hambleton, R.K., Swaminathan, J., Algina, J. & Coulson, D.B.: 1978, 'Criterion-referenced testing and measurements: A review of technical issues and developments', *Review of Educational Research*, **48**, 467–510.

Isaacs, I.: 1987, 'Using problem solving teaching styles to prepare candidate for the CXC Basic Proficiency Examinations', *Educational Studies in Mathematics*, **18**, 177–180

Linn, R.L.: 1983, 'Testing and instruction. Links and distinctions', *Journal of Educational Measurement*, **20**, 179–189.

Messick, S.: 1988, 'Validity', in Linn, R.L. (Ed.), *Educational Measurement* 3rd edition, Macmillan Publishing Co., New York.

Nattall, D.L.. 1987, 'The validity of assessments', *European Journal of Psychology in Education*, **2**, 109–118.

Rovinelli, R.J. & Hambleton, R.K.. 1977, 'On the use of content specialists in the assessment of criterion-referenced test item validity', *Dutch Journal of Educational Research*, **2**, 49–60.

Snow, R.E.. 1980, 'Aptitude and achievement', in Schrader, W.B. (Ed.), *New directions for testing and measurement Measurement Achievement: Progress over a decade-proceedings of the 1979 ETS Invitational Conference*, Jossey-Bass, San Francisco.

Desmond R. Broomes
University of the West Indies,
Barbados

James A. Halliday
Caribbean Examinations Council,
Barbados

MURAD JURDAK

ASSESSMENT IN MATHEMATICS EDUCATION IN THE ARAB COUNTRIES

1. INTRODUCTION AND BACKGROUND

The Arab region includes 21 countries, extends over a vast area of two continents (Asia and Africa), and currently has a total population of 220 million. Consequently, statements about Arab countries must be couched in generalities if they are to be true. Fortunately, the commonalities in the educational assessment systems of the Arab countries outnumber the idiosyncracies of the individual countries. In this paper, the commonalities in the educational assessment systems will be described in as specific terms as possible.

It is only in the last five decades that the Arab countries have emerged as independent sovereign states. The concerns of the emerging state focused, in the early years of independence, on three issues: consolidation of *state authority* over education, provision for *mass education*, and *Arabization* of education (Bashshur, 1982) The consolidation of education was reflected in the State's assuming responsibility for education (30 percent of students in the Arab countries were in private schools in 1950–51; this percentage now is negligible in all Arab countries with the exception of Lebanon) and in the formulation of national educational policies, laws, and institutions. In assessment, consolidation was conspicuous when external government examinations and diplomas were instituted as substitutes for the exams and diplomas of the colonial powers. The provision for mass education was caused by the surge of students who joined schools as the fledgling states increased opportunities for free education tremendously. For example, between 1960 and 1985 the number of students more than tripled in the elementary schools and increased by nine times in the secondary schools (Sara, 1990). As far as assessment is concerned, the surge of students dramatically increased the strains on the already rigid systems of education, resulting in the mandating of Arabic as the language of instruction but also in the increased coordination among Arab states. Arabization culminated in 1964 with the establishment of the Arab League Education, Culture, and Science Organization (ALECSO).

As they were coping with the pains of growth, the Arab countries had to respond to new challenges posed by outside or inside forces. Three

M. Niss (ed.), Cases of Assessment in Mathematics Education, 35–41.

challenges were universally recognized by the conferences of the Ministers of Education and those responsible for economic planning (Unesco, 1966, 1970 & 1977): orienting education towards socio-economic development, introducing and adapting technology, and improving the quality of education. These challenges were reflected in the accelerated efforts to formulate long-term educational plans, improve curricula and learning environment, and improve teacher training. However, assessment systems remained unchanged.

The colonial powers, mainly Britain and France, left behind educational systems modelled, more or less, after their own. These educational systems featured curricula organized in strict hierarchical levels (grades or classes) grouped into definite states: elementary (6 years), intermediate (3 years), and secondary (3 years). In addition to variations in the number of years in each stage, some Arab countries adopted a two-stage system: basic education (8 or 9 years) and secondary education (3 years). The progress of students through the system was — and is — determined by the results of assessment for promotion at the end of each level (grade or class) to the next immediate one or, at the end of each stage, for promotion to the next stage.

2. MATHEMATICS AND ASSESSMENT

Mathematics has always enjoyed a special status in the curricula in the Arab countries for two reasons. First, mathematics is viewed as a universal culture-free subject that does not threaten the Arab-Islamic culture. Second, mathematics is viewed as the basis for science and technology — highly valued by the Arab countries. For these reasons, mathematics was chosen as the first area for an inter-Arab cooperative curriculum project (Jurdak & Jacobsen, 1981).

Assessment in mathematics is closely related to the demands of the educational system and the dominant conception of the nature of mathematics (Figure 1). In the Arab countries, the prevailing conception of mathematics is that it is a neutral and external body of knowledge that has an inherent hierarchical structure. As such, the content of mathematics consists of labels and symbols, facts, principle, and algorithms. Assessment observes and measures the recall and recognition of labels, symbols, and facts; skills in performing algorithms; and, solving "typical problems". Implicit in assessment are the different kinds of behaviors of different cognitive levels. However, the content-by-behavior matrix is rarely used explicitly as an organizing construct in assessment. In the conception reflected in this matrix, the content of mathematics can be classified into mutually exclusive strands (whole numbers, rational numbers, ...) each of which is unidimensional and cumulative. The merging of this concept of mathematics with the structure of the educational system results in a

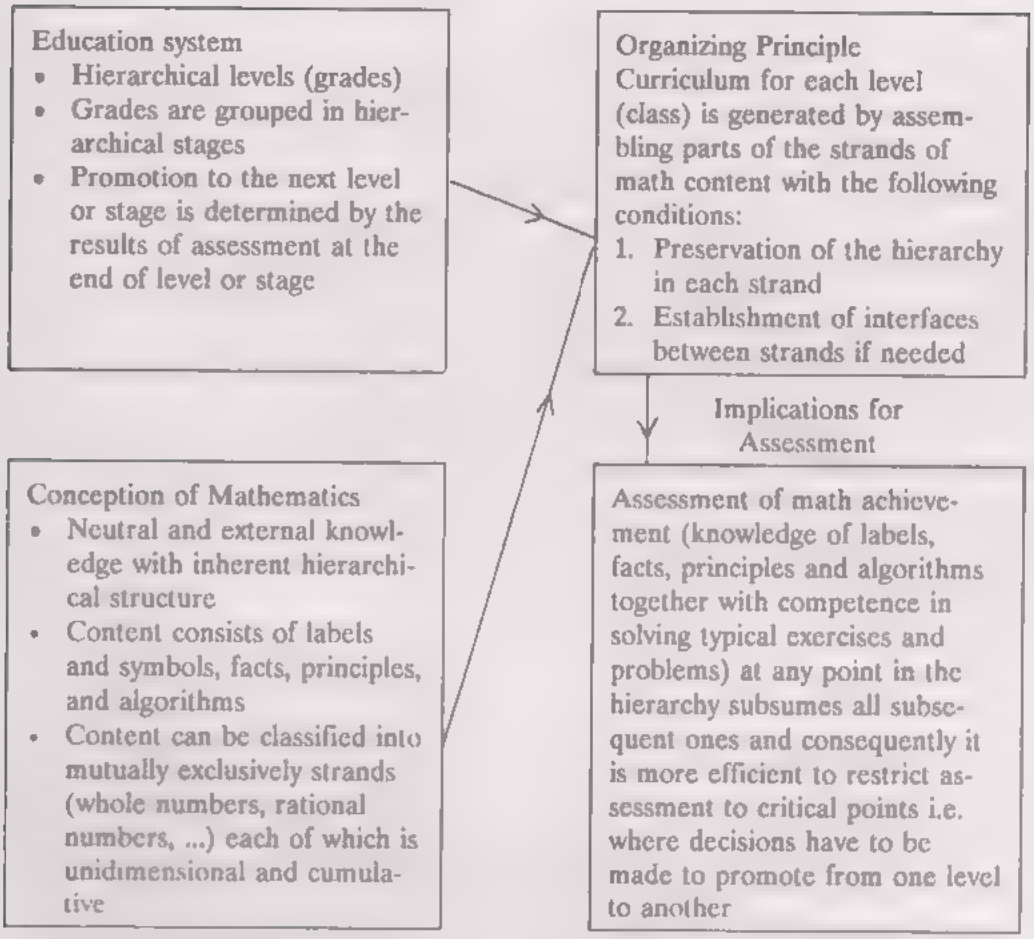


Figure 1 *Typical relationships between the education system, the prevalent conception of mathematics, and assessment in mathematics in the Arab countries.*

simple, yet powerful, organizing principle: The curriculum for each level (class) is generated by assembling parts of the strands of mathematical content, preserving the hierarchy in each strand, and, if needed, establishing interfaces between strands. One implication of this organization is that the assessment of achievement at any point in the hierarchy of a strand assumes the assessment of the knowledge of all content in that strand below that particular point. Thus it is more efficient to restrict the assessment of students to critical points, i.e., where decisions have to be made to promote from one level to another, or from one stage to another.

In general, variations of the three types of assessment can be identified in all of the Arab countries. The first type is done by the school for its students (henceforth called *internal assessment*) and the second is done by the State for all students at the end of each stage (henceforth called *external assessment*). A third type of assessment is more covert and not planned explicitly (henceforth called *hidden assessment*).

Internal Assessment

Invariably, the primary purpose of the internal assessment of mathematical performance is to select students for promotion to the next level (class). A secondary purpose is to monitor and control the learning of students during the academic year. With the exception of the scheduling of exams, the teacher almost exclusively initiates and controls the process of internal assessment. The teacher has the power to determine what information is to be gathered, the nature of the mathematical tasks used in assessment, the scoring procedures, and the collecting and recording of information. Basically, information is gathered on the individual student. Most often, the assessment instruments are teacher-constructed tests which consist of written mathematical tasks involving skills and problems typical of the problems solved during the semester or the year. Figure 2 includes some examples of assessment tasks from some Arab countries (translated from Arabic).

I. a) Complete: $\frac{5}{7}=\frac{\square}{21}=\frac{30}{\square}=\frac{\square}{63}$.

b) Simplify: $\frac{90}{40}, \frac{36}{54}, \frac{45}{165}$.

c) Ahmad has 50 riyals. He bought a pair of shoes for 27 3/4 riyals and a bag for 12 7/8 riyals. How many riyals are left with him?

(Yemen, fourth elementary (Grade 4), Fractions)

II. A and B are two points on a circle whose center is O such that $A\hat{O}B=50°$. X is a point on the major arc AB.

a) Prove that $A\hat{X}B=25°$ whatever the position of X on the major arc AB.

b) What is the measure of $A\hat{Y}B$ where Y is a point on the minor arc AB

(Saudi Arabia, third intermediate (Grade 9), Geometry)

III. a) What is the value of a if the straight line $y-ax+3=0$ passes through the intersection of $y+x+2=0$ and $2x+3y+5=0$?

b) The points $A(1,1)$, $B(4,5)$, $C(9,5)$, $D(6,1)$ are vertices of a quadrilateral whose diagonals are AC and BD. Prove that the two diagonals are perpendicular. Find the area of the quadrilateral.

(Egypt, second secondary (Grade 11), Coordinate Geometry)

Figure 2 *Examples of assessment tasks in mathematics from some Arab countries*

These instruments are criterion-referenced, the criterion being a certain score for success (most often 50 percent) on an absolute scoring scale (most often the percentage scale). The assessment instruments normally suffer

from many psychometric deficiencies including the lack of content *validity* (content selection for assessment purposes is almost always left to the teacher), *reliability* (the scoring and scoring procedures are subjective), and *comprehensiveness* (a narrow range of content and cognitive processes is tested). Standardized tests (and hence norms), even in the very few cases where they exists, are not used by teachers in internal assessment.

External Assessment

External assessment is done by the State (Ministry of Education) at the conclusion of each stage for the purpose of selecting those students who will be promoted to the next stage. External assessment is controlled by the Ministries of Education in terms of decisions concerning the nature of mathematical tasks to be used for assessment, scheduling and timing of the exam, and its administration, scoring, coding, and reporting. Operationally, the process is handled either by the staff of the "examination section" in the Ministry of Education or by a committee of mathematics teachers appointed by the Ministry. The tests used are unified (for the district or for the country) and they consist of written items modelled after the "typical" exercises and problems studied earlier. Because mathematics is conceived of as cumulative and unidimensional, test items are normally taken from the mathematical content of the last level (class) of the stage rather than from all levels (classes) of the stage. Tests used in external assessment tend to include more mathematical tasks with multiple parts which are not scorable independently thus affecting negatively the reliability of the tests. Since instruments used for external assessment are prepared and scored by a team rather than an individual teacher, they tend to be superior, in terms of psychometric properties, to the instruments used in internal assessment. Scoring is again on an absolute scale and success is determined by the total score on all subjects.

Of special significance is the external examination at the end of secondary school. The scores in the external assessment determine not only whether a student is admitted to the university but also what major subject the student is allowed to follow. Normally the score in mathematics carries a lot of weight for admission to the professions and sciences.

Hidden Assessment

Hidden assessment refers to an underlying or covert system that modifies the declared objectives, processes, and products of the original system. In the Arab countries, hidden systems of assessment of this kind serve a dual purpose: First, the hidden system provides incidental information for modifying the elements of the system, and second, it acts as a stabilizer to the educational system.

As mentioned earlier, the main purpose of assessment in the Arab countries is the promotion of students to the next level or stage. The hidden system inadvertently modifies the purposes of teaching mathematics, the curriculum, and teaching methods. Whatever the declared purpose of teaching mathematics, the hidden assessment transforms it into a simple, realistic yet meaningful purpose, i.e., upward mobility in the educational system. To achieve this purpose and in the absence of norms and well-defined criteria, the single most important criterion becomes the coverage of the elements of mathematical content typically included in tests and exams. What is emphasized or deleted in actual teaching depends, to a large extent, on the probability of its being sampled in the tests. Teaching methods also serve the requirements of the assessment system by employing instructional strategies that promote recall, recognition, knowledge of algorithms, and problem solving of typical problems. Because of the special status and weight given to mathematics, the hidden system has aggravated these problems in mathematics education.

In the absence of norms or criteria (other than content coverage), the critical points in the assessment system (end points of levels or stages) act like safety "valves" to achieve balance in the educational system. If the flow in the system is more (or less) than expected or desired, valves are adjusted to increase (or decrease) the flow. Some examples will illustrate this dynamic relationship. Many of the educational systems in the Arab countries have suffered from chronic problems of drop-outs and repetitions. As a solution, many countries have adopted the "automatic promotion" policy, i.e., all students are automatically promoted from one level to another irrespective of their performance. A second example is the abolishment of the external assessment at the end of the elementary stage in order to extend compulsory education from six to nine years. Through this process, the hidden assessment acts like a mechanism to perpetuate the present educational system by neutralizing changes and innovations.

3. NEW TRENDS

In recent years, indications of new developments in educational assessment are discernible in some Arab countries. First, there is a trend towards expanding the scope of the purposes for assessment and evaluation. In some countries (Kuwait and Bahrain) national projects of comprehensive evaluation have been started (Sara, 1990) Such projects follow the system approach and are intended to provide information for a wide variety of decisions including curricula and teaching. Second, there is a trend to develop and use a variety of assessment tasks and instruments. Egypt is developing a national item bank to be used not only for external assessment but also for diagnostic and formative assessment purposes by the teachers (Srour, 1989). In Bahrain, information for the assessment of

students is collected during the course from sources other than written tasks (projects, papers, observation). Third, there is a strong call for introducing basic changes in the structure of educational systems moving them toward more flexibility. A four-year multidimensional regional research project on the status and future of education in the Arab countries is noted here. The final project report (Ibrahim, 1991) recommends, among other things, that the strict hierarchical closed structure of the "educational ladder" be replaced with the open and flexible concept of the "educational tree". Such basic changes in the goals, structure, and content of curricula call for aligning the assessment system with such changes.

REFERENCES

Bashshur, M. 1982, *Trends in Arab Education* (2nd Ed.) (in Arabic), Arabic League Education, Culture and Science Organization (ALECSO), Tunis.

Ibrahim, S.: 1991, *Teaching the Arab nation in the twenty-first century* (in Arabic), Muntada Al-Fikr Al'Arabi, Amman.

Jurdak, M. & Jacobsen, E. 1981, 'The evolution of mathematics curricula in the Arab States', in Morris, R. (Ed.), *Studies in Mathematics Education Vol. 2*, UNESCO, Paris.

Sara, N. 1990, *Arab education since 1950: Achievements, problems, challenges* (in Arabic), Muntada Al-Fikr Al'Arabi, Amman.

Srour, A. 1989, *Developing education in Egypt* (in Arabic). Ministry of Education, Cairo.

UNESCO. 1966, *Final report of the conferences of ministers of education and those responsible for economic planning in the Arab States*, UNESCO, Tripoli.

UNESCO. 1970, *Final report of the conferences of ministers of education and those responsible for economic planning in the Arab States*, UNESCO, Marrakesh.

UNESCO. 1977, *Final report of the conferences of ministers of education and those responsible for economic planning in the Arab States*, UNESCO, Abu Dhabi.

Murad Jurdak
Division of Educational Programs,
American University of Beirut,
Lebanon

JOHN A. DOSSEY & JANE O. SWAFFORD

ISSUES IN MATHEMATICS ASSESSMENT IN THE UNITED STATES

1. INTRODUCTION

The results of comparative studies of United States students' mathematical achievement (McKnight et al., 1987; Crosswhite et al., 1987; Dossey et al., 1988) in national and international arenas have heightened the United States' awareness of the importance of mathematics as a tool for stability in an era of rapid change (MSEB, 1989a, 1989b; Adelman & Alsalam, 1988; Johnson & Packard, 1987). At the same time, mathematics educators and others set out to describe how the United States' society might measure both the growth of mathematical ability in individuals and in the society itself (Alexander & James, 1988; Kulm, 1990; Raizen & Jones, 1985; Resnick, 1987; NCTM, 1989). These attempts also brought with them a number of reports dealing with the dangers of testing and the role testing might play in stratifying society (Nationa. Commission on Test and Public Policy, 1990).

This national obsession with assessment as a means of measuring the progress of education and the future productivity of society has resulted in a massive system of loosely-connected sets of indicators of the nation's level of mathematical literacy. These range from the National Assessment of Educational Progress, to assessment programs in individual states, to standardized testing in schools, to college entrance examinations, and to national plans for assessment systems that will correctly inform the nation about the levels of mathematical achievement and potential of its citizens.

2. NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS

Since the late 1960s, the major barometer of the current status of the mathematical health of the United States' students has been the mathematics assessment of the *National Assessment of Educational Progress* (NAEP), a program of the United States Department of Education. Since its inception, NAEP has gathered information on trends in student achievement in eleven different academic areas. Mathematics assessments have been carried out in 1973, 1976, 1978, 1982, 1986, and 1990. The analyses

M. Niss (ed.), Cases of Assessment in Mathematics Education, 43–57.

of these assessments have provided a basis for a number of reports on the health of mathematics education in the United States (Lindquist, 1989).

In 1988, the United States Congress, reacting to public concern over the state of education, increased the emphasis to be given to mathematics in the NAEP program by requiring assessments in mathematics on a biennial basis beginning in 1990. In addition, the Congressional action added a new dimension to the NAEP assessments. It provided for separate voluntary, state assessments in the years of 1990 and 1992. This provision makes possible state-to-state and state-to-nation comparisons of student achievement (Mullis, 1990).

To establish a set of objectives for this broader use of the NAEP assessment, the National Assessment Governing Board (NAGB) worked through the Council of Chief State School Officers (the elected or appointed leaders of the educational departments in each of the 50 states) to establish a new set of mathematical objectives for the 1990 NAEP mathematics assessment (NAEP, 1990). These objectives are based on the expectations set forth in the National Council of Teachers of Mathematics's *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989).

The items of the 1990 NAEP mathematics assessment reflected a balance of items drawn from the five content domains of *Numbers and Operations*; *Measurement*; *Geometry*; *Data Analysis, Statistics*, and *Probability*; and *Algebra and Functions*. These five content domains were crossed with a three level model of mathematical abilities which contained the categories of *Conceptual Understanding*, *Procedural Knowledge*, and *Problem Solving*. The items are a mixture of approximately 4/7 multiple choice and 3/7 open-response items. In addition, approximately 3/7 of the items at a given grade level may be answered with the use of a calculator.

These items showed a great deal of variability in the types of tasks they required of students. They ranged from fairly direct questions such as the following items on percentage:

Which of the following is true about 87% of 10?

A It is greater than 10.
B It is less than 10.
C It is equal to 10.
D Can't tell.
E I don't know.

which only 75 percent of 12th Graders could answer correctly, to more complex open-response items such as the room arrangement problem shown below (Figure 1). Sixty-seven percent of the 8th Graders taking this item completed the task successfully. As the NAEP examination is more of a broad census of mathematical abilities, it lacks deep probes into the

depth of student's mathematical knowledge. One of the more difficult algebra items on the 1990 assessment was:

Solve for x in the equation below.

$$(x+1)^2-3(x+1)=-2$$

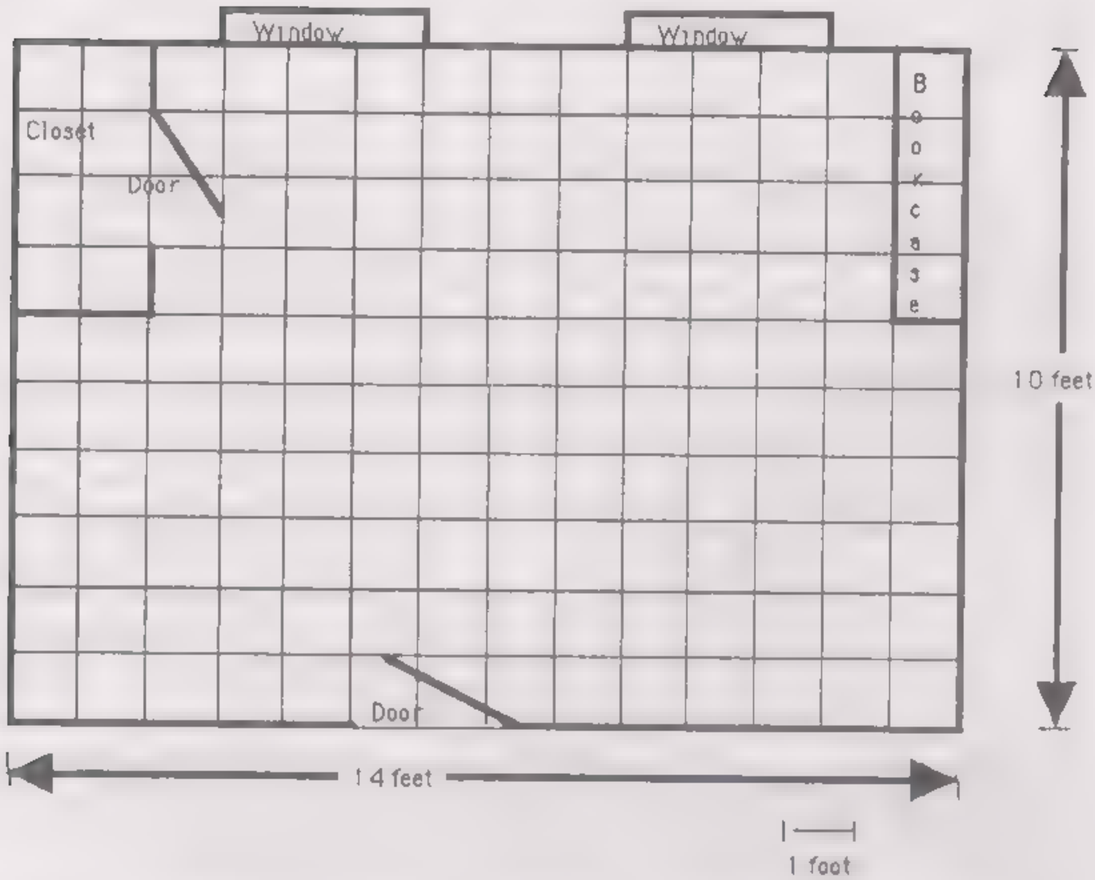
This problem, cast in a form not typically seen in an algebra classroom, was correctly completed by only 11 percent of the students. Another 18 found one of the roots to the equation, but not both. This type of item shows the lack of students' ability to make small steps from what they have practiced in the classroom to new, but slight extensions, of the material they have had a strong opportunity to learn (Mullis et al., 1991).

The trial state-by-state program, in which 40 states participated in 1990, extends the influence of the NAEP goals for mathematics from those participating in the national sample to students in a majority of the states. In doing so, the NAEP mathematics assessment moves closer to becoming a "national mathematics test" for school children in the United States (Hambleton, 1990). Over the years, NAEP has served as one common measure of national progress in mathematics, as it was the only mathematics examination given to a randomly drawn national sample of United States' youth. This new emphasis on the use of the NAEP tests has already created questions about the NAEP process. Now with state-by-state comparisons, the stakes have become high and children are being prepared for the NAEP assessment in some schools. Thus, the nature of the testing situation and the context for interpreting the results must change.

Another question that has arisen is the relation between the trend lines for student performance at 9, 13, and 17 years-of-age and future student achievement. The NAEP assessment has always been a mark of student achievement, unlike college entrance examinations which assess ability. Further, the NAEP assessment, while broad in nature, does not cover all forms of outcomes or even content taught in mathematics classes. It is a cross-sectional survey of student achievement. Critics question whether the increase in stakes associated with the use of NAEP scores for state-by-state comparisons will narrow the curriculum taught to only the topics which NAEP covers.

Some have argued that the NAEP testing should be taken to the student and building level. Such a use would extend the assessment beyond its design. At present it employs a balanced-incomplete-block design with several different blocks per grade level. Using the block design, the NAEP assessment has included far more items than standardized tests.

The very nature of the way in which the present NAEP examinations are configured makes them well suited for finding differences in the performance of students from different states, but it can do little to explain the sources of those differences. There is nothing in the information gathering process with the NAEP assessment that provides information on the student's opportunity to learn the material tested. No information is



The diagram above represents a scale drawing of John's room. Each side of a block in the diagram represents 1 foot. John has four pieces of furniture that he needs to put in the room.

The measurement of the furniture are:

bed	6 feet long,	3 feet wide
desk	5 feet long,	3 feet wide
chest	5 feet long,	2 feet wide
bookcase	5 feet long,	1 foot wide (already in place)

In arranging the furniture, John must follow these rules:

- o The doors may not be blocked
- o Each piece of furniture must have at least one side against a wall of the room
- o The chest is too tall to be placed against a window.

The bookcase has already been put in place. On the diagram a scale drawing of the bookcase shows where it has been put. Decide on a way that John could arrange the other three pieces of furniture so that the total arrangement follows all the rules. On the diagram, show that arrangement by drawing in each piece of furniture in its place. Draw each one to scale, using the same scale as was used to make the diagram. Label each piece of furniture.

Figure 1 *NAEP item*

gathered on local or state curricula and the nature of their implementation.

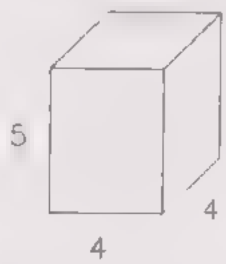
3. ASSESSMENT PROGRAMS IN INDIVIDUAL STATES

As a result of the calls for reform in school mathematics and an overall move to make schools more accountable, many individual states in the United States instituted forms of statewide assessment in mathematics during the 1970s. These took a variety of forms from clones of standardized achievement tests to carefully thought out programs of assessment of individual skills. In this section we will look at state assessment programs in several states to illustrate some of the innovations that are underway in the United States.

One issue in state assessment is the population assessed. In some states all students participate in the testing program, while in other states a sample of students is selected as participants. This leads to differences in both the view of the testing program as a high- or low-stakes venture on the part of the schools. This issue is related to the manner in which scores are reported at the state, school, or individual level and the consequences associated with these scores. Other issues affecting the nature of state assessments are the form of tests given — multiple-choice, open-response, or a mixture of the two. This facet is accompanied by the question of whether students are allowed to use calculators in the completion of all or a portion of the test.

The state of Massachusetts established a statewide assessment program with two purposes: to compare the effectiveness of public schools and to give guidance for the improvement of curriculum and instruction. To serve the dual intent, three major components have been developed. The first component is a multiple-choice test administered in alternate years to all Grades 4, 8, and 12 students in the state. The results of this component are used to calculate school scores for the purpose of school comparisons. The second component is an open-ended test given in reading, mathematics, science, and social studies. These questions are administered as part of the multiple-choice assessment but do not form part of the school reports. In 1989, the third component, a performance assessment was added. The performance tasks involves the application of mathematical and scientific concepts to solve problems. The tests are administered in alternate years to the written assessment of Grades 4 and 8 students working in pairs. In 1989, over 2,000 pairs of students were assessed. The performance tests are scored by trained volunteer teachers and curriculum coordinators and the results of both the open-ended and performance tests reported with instructional implications in order to help principals, teachers, and curriculum coordinators improve curriculum and instruction (*Massachusetts Education Assessment Program*, 1989, 1990). California also introduced open-ended questions into their Grade 12 assessment test in 1987–88 but only had funds to score 2,500 of the total of 240,000 responses statewide. Nevertheless, open-ended questions, such as the item in Figure 2 below, are expected to become a regular part of the mathematics portion of the

California assessment program at all grade levels in the future (*California Assessment Program*, 1989; Pandey, 1991).



You have been asked to wrap the box above for a birthday present. The best dimensions for a single sheet of wrapping, allowing for overlap, are

A: 5 by 16
B: 6 by 14
C: 9 by 18
D: 10 by 20

Figure 2 *California Assessment Program item*

Connecticut also plans to use performance tests in their high school mathematics and science assessment.

The states of Michigan, Connecticut, and Missouri have all adapted their state assessments to allow the use of the calculator on the assessments. Connecticut, as a state, purchased calculators for students to use in their classes, as well as on the *Connecticut Mastery Test*. Michigan and Missouri each allow students to use their own calculators in completing the mathematics portion of the state assessments. Other states, Illinois, Texas, and California are moving toward the integration of calculators into their assessments of student mathematics.

The state of Illinois has worked to develop a system of state learner outcomes in mathematics that provides for reporting student achievement at Grades 3, 6, 8, and 11 on a broad curricular front. Scale scores are reported at the school level for grades 3, 6, 8, and 11 in the area of mathematics. There is an overall core score and scale subscores in six areas, allowing for the comparison of schools on subscale areas and downplaying the use of a single scale score for the school as the point of comparison.

The existence of state assessments in mathematics has a strong influence on school mathematics curricula at the state and local levels. This influence can, as in the NAEP setting, have a propaedeutic effect or a narrowing effect on the curriculum. Vermont has addressed this issue with the establishment of a portfolio form of assessment for state purposes at Grades 4 and 8. This program turns away from testing as the sole vehicle for establishing student progress at the state level. The individual student portfolios provide a careful look at the growth of individuals across a

school year and between levels of their schooling. Its focus is on good work in problem solving and individual production of work in mathematics, not on speed and accuracy under time pressure. At present the Vermont experience is a unique point in state assessments, but many other states are considering moves in this direction.

4. STANDARDIZED TESTS

Standardized tests are widely used in United States elementary and secondary schools. In fact, it has been estimated that 200 million standardized achievement tests are given in the United States each year (Mehrens & Lehmann, 1987). By a standardized test we mean a commercially prepared test that provides items for obtaining samples of students' behavior under uniform procedures. Usually a standardized test has been administered to a norming group so that a student's performance can be interpreted by comparing it to the performance of others, the norm-referenced group.

Examples of standardized achievement tests used in the United States include the *Stanford Achievement Test*, published by the Psychological Corporation of San Antonio, Texas, which is one of the most widely available achievement batteries for assessing school achievement in Grades 1 through 9 in the United States. It contains a number of subtests, including three in mathematics: concepts of number, mathematics computation, and mathematics applications. Another example of a multilevel comprehensive test battery is the *Iowa Tests of Basic Skills*, published by Riverside Press of Chicago, Illinois. It consists of 11 separate subtests measuring skills in five areas including mathematics skills. The *California Achievement Test* (CAT), published by CTB/McGraw-Hill of Monterey, California, is another popular standardized achievement battery in the United States. The latest versions cover the traditional verbal and quantitative topics found in Grades K–12. The quantitative topics include computation, concepts, and applications. All of the U.S. standardized tests use a multiple-choice format for their mathematics items. However, the most recent version of the California Achievement Test has an increased emphasis on problem-solving skills at the upper grade levels with items presented in such a way as not to require computation of a numerical answer.

Several issues surround the use of standardized tests. One is test content. Most publishers of standardized tests present a very complete list of the topic areas covered by their tests. However, critics argue that the items test primarily recall of facts and the execution of routine procedures. A study examining six standardized mathematics tests for Grade 8 (Romberg, Wilson & Khaketla, in press) found on average that 89 percent of the items tested procedural knowledge in contrast to conceptual knowledge. Many

teachers perceive that tests measure primarily low level cognitive skills and gear their curriculum accordingly.

Since standardized tests play such an important role in United States schools, researchers have investigated the alignment of standardized tests with textbook content (Freeman et al., 1983). Their research revealed that textbooks emphasized computation far more than did any of the standardized tests but that a large proportion of the material in textbooks was not covered on standardized tests. Further, the match between texts and tests varies widely with different test-text pairs.

Standardized tests have also been examined to determine whether or not they are appropriate instruments for assessing the content, process, and levels of thinking called for in the NCTM *Standards* (NCTM, 1989). A study of the six most widely used standardized tests at the state and district levels in schools in the United States found the relationship of the tests to the Standards to be generally weak in six of the seven content areas with most of the items belonging to the content area of numbers and number relationships (Romberg et al., in press). Further, only one of the six process areas, computation/estimation, was extensively covered by the tests.

Also at issue is the use or non-use of calculators on standardized tests. Only the most recent edition of the Stanford Achievement Test gives calculator norms. Because of the impact of standardized tests on the curriculum, critics claim that the inability to use calculators on them has inhibited calculator use in the classroom and has encouraged the retention of obsolete skills in the curriculum. One study found that about 25 percent of the teachers reported that they decreased emphasis on calculator activities because they are not allowed on standardized tests (Romberg, Zarinnia & Williams, 1989).

Another issue is the use of standardized tests, which are many and varied. Originally used as instructional aids, standardized tests are now used to evaluate academic progress; to determine developmental levels of students; to diagnose students' specific strengths and weaknesses; to select, classify and place students; to diagnose group strengths and weaknesses; to compare alternative instructional procedures; and to serve as a dependent variable in educational research. Generally, no validity studies for most of these uses have been conducted. Hence it is possible that the decisions made using standardized tests scores might be made on the basis of what is essentially measurement error (Airasian, 1985).

In addition to instructional uses, standardized tests have been increasingly used in the United States as *accountability* measures: School boards, state legislators, and the public press use standardized test scores to assess local school improvements as well as overall school quality, teacher and administrator competency, and program effectiveness. In most states scores are published in the newspapers on a school-by-school or district-by-district basis and in some locales affect real estate prices. In some states, even teachers' advancement or merit pay have been tied to the standardized test

scores of their students. As the stakes rise, there is increased pressure to see that scores do likewise. Cannell (1990) provides a carefully constructed analysis of the use of norm-referenced-testing as a basis for the determination of accountability and student progress. He finds the process lacking from a number of viewpoints with widespread *teaching to the tests* and even cheating. Whatever the means, standardized test scores have risen until almost all of the states report that a majority of their students are scoring above the national norm. These results viewed locally are a source of pride, but viewed nationally bring into question the validity of the norms, the norming process, and the use of standardized tests to assess teacher and school effectiveness.

5. COLLEGE ENTRANCE EXAMINATIONS

A special form of standardized tests is *college entrance examinations*. The United States does not have a universal examination at the end of secondary school to assess what students have learned during their schooling or to assign levels to students' achievement. Instead of *school exit exams*, several million United States students annually voluntarily take one of two standardized college entrance examinations. Most colleges and universities in the United States require scores from one of these exams to supplement an applicant's secondary school academic record in the admission decision. Since secondary schools vary considerably in the United States, such entrance examinations are used to provide a common yardstick with which colleges can compare the abilities of applicants coming from different backgrounds and educational systems (Educational Research Service, 1981).

The oldest of these examinations is the *Scholastic Aptitude Test* (SAT) produced by Educational Testing Service (ETS) for the College Entrance Examination Board, the College Board for short. The SAT is taken by almost 1.5 million 11th and 12th Grade students annually. In 1989–90, it is estimated that 40 percent of the high school graduating class in the United States took the SAT test (Dodge, 1990).

The SAT is a three-hour, multiple-choice test to "measure the verbal and mathematical abilities developed over many years, both in and out of school" (College Entrance Examination Board, 1989). At present, the mathematics questions, covered in two 30-minute sections, test students' ability to solve problems involving arithmetic, elementary algebra, and geometry. Students receive two scores, verbal and mathematics, each reported on a scale of 200 to 800. Each year a different form of the test is given. The SAT has recently been revised to reflect changes in the educational reform movement. Beginning in 1994, it will consist of two portions, the *SAT-I*: Reasoning Tests and *SAT-II*: Subject Tests. SAT-I will consist of revamped, enlarged versions of the verbal and mathematics tests,

but will include added emphasis on critical reading and student generated responses in mathematics. In addition, the mathematics tests will reflect increased emphasis on data analysis and applications and students will be allowed to use a hand calculator on the examination. Some examples of items reflecting differing levels of potential of calculator usage from the SAT-II are as follows:

Calculator inactive:

If $f(x)=2$ and $f(g(x))=-x$, then $g(x)=$

(a) $-3x$

(b) $-\frac{x}{2}$

(c) $\frac{x}{2}$

(d) $2-\frac{x}{2}$

(e) x

Calculator neutral:

If $2x^2+4x=3$, then, to the nearest tenth, what is the positive value of x?

(a) 0.6
(b) 0.8
(c) 1.2
(d) 1.6
(e) There is no positive value of x that satisfies the equation.

Calculator active:

What is the area of a right triangle with an angle of 28 and with longer leg of length 13?

(a) 40
(b) 45
(c) 75
(d) 90
(e) 159

SAT-II will consist of a major new test in writing, subject matter tests, and other tests designed to be useful in placement or in determining a student's command of the English language. The new versions of the SAT-I will be first used in the spring of 1994. The changes in SAT-II will become available over the period from 1991 to 1994 (College Entrance Examination Board, 1990).

The second college entrance examination used in the United States is the *American College Testing* (ACT) program. The ACT test attempts to assess the general educational development of high school students and their ability to perform college-level work (American College Testing Program, 1977). The ACT also provides a personal interest inventory and self-descriptive and self-evaluation information to be used for course and career placement. The ACT is a battery of four academic tests which in 1989 was revised to emphasize a wider range of mathematical knowledge,

more abstract reading skills, and how well students deal with scientific concepts. The scores are reported as standard scores with a range of 1 to 36 and a composite score which is the average of the four standard scores. The new expanded ACT allows for the reporting of subscores for the purpose of course placement in college. In mathematics there are three subscores: pre-algebraic/elementary algebra, intermediate algebra/coordinate geometry, and plane geometry/trigonometry. More than 800,000 students take the ACT annually.

A number of issues surround the use of college entrance examinations. One is the importance placed on scores. The exams purport to be aptitude rather than achievement tests. However, the general public views them as measures of the effectiveness of American schools and uses them as barometers of United States education. Each year, front page headlines announced the newest average scores and compare them with the previous years' scores. SAT scores reached a peak in 1963 and then fell into a deep decline through the seventies. The decline triggered a variety of reports and reactions and a discussion of the validity of such scores as measurement of educational quality.

Another issue is the validity of college entrance examinations. The examinations are designed to assist admissions officers in making admissions decisions. Some feel that using a standard test makes the college admissions process a fairer process because is provides a common yardstick across different geographic areas and socioeconomic levels. Critics argue that such tests tend to maintain the status quo rather than promote reform. Because the SAT is the older of the two tests and used by the more selective private colleges, more critical scrutiny has been focused on it than on the ACT, although many of the points made by critics also apply to the ACT.

One measure of the validity of an admissions decision is freshman grades. Validity studies of the SAT produce correlation coefficients of SAT scores with freshman grade-point average that fall in the 0.4 to 0.5 range (Wiersam & Jurs, 1990, p. 362). Crouse and Trusheim (1988) in *The Case Against the SAT* argue that their research shows that for most colleges, an admissions policy based solely on applicants' high school records would admit and reject nearly the same students as one which uses both SAT and high school records. They conclude that the SAT is a costly redundancy (Crouse, 1986). ETS responds that in their studies, correlations increased by 0.07 to 0.10 when SAT scores were combined with high school records, a 15 to 18 percent improvement (Cameron, 1989; Willingham & Ramist, 1982).

Test bias is another issue that shrouds the use of college entrance examinations. Historically, men have scored higher than women on both the SAT and ACT and minorities (blacks, Hispanics and American Indians) have scored lower than whites. Minority women also score lower than men in their own ethnic group (College Entrance Examination Board, 1988).

Critics claim that college entrance examinations, particularly the SAT, are geared toward white males, leaving women and minority-group students at a disadvantage. A number of research studies have looked critically at sex difference on the SAT and found support for the claims of bias (Burton, Lewis & Robertson, 1988; Rosser, 1989; Wilder & Powell, 1989). Others claim that test scores are possibly influenced by coaching and that this makes scores open to undue influence related to family's wealth. Whether or not coaching improves scores is a hotly debated issue (College Entrance Examination Board, 1989; Smyth, 1989; Wilson, 1990). Nevertheless, critics claim that the SAT is biased in content, context, validity and use while the College Board counters that factors such as differences in population size, academic background, and socioeconomic status explain much of the differences in mean scores.

6. CLASSROOM ASSESSMENT AND THE *STANDARDS*

Classroom assessment takes a variety of forms in the United States but is dominated by teacher-made tests of procedural knowledge. Most textbooks now provide teachers with tests to accompany each chapter. Some provide pre-chapter as well as post-chapter tests or offer computer grading programs. Teachers, however, continue to develop their own tests or to modify publishers' tests to reflect their own instructional emphasis.

Recognizing that tests often drive the curriculum, the National Council of Teachers of Mathematics included a working group on evaluation when they established the Commission on Standards for School Mathematics in 1987. The work of the Commission culminated in the publication in 1989 of the *Curriculum and Evaluation Standards for Schools Mathematics* (National Council of Teachers of Mathematics, 1989). As can be seen from the title, the document includes standards for evaluation. The Evaluation Standards propose changes in the processes and methods of assessment. Key to this is the notion that student assessment should be an integral part of teaching and should be based on multiple assessment methods, including written, oral, and demonstration formats which use calculators, computers, and manipulatives. All aspects of mathematical knowledge and its connections should be assessed. The focus should be on a broad range of mathematical tasks rather than on a large number of specific and isolated skills organized by a content-behavior matrix.

The Standards view evaluation as a tool for implementing and effecting change. The position taken in the Standards calls for radical change in classroom assessment in the United States. Tests, both teacher-made and standardized, need to change; but the Evaluation Standards call for changes beyond the mere modification of tests. Their main purpose is to help teachers better understand what students know and make meaningful instructional decisions.

7. CONCLUSION

The foregoing discussion of issues facing the assessment of student progress in mathematics in schools in the United States shows both the significant effects that assessment has on student achievement in mathematics and the myriad of problems associated with attempts to evaluate student achievement. Educational measurement and its application to school settings has reached an almost crisis stage in the United States. Assessment is seen as both a tool to further reform, while at the same time an impediment to change. The resolution of this paradox may hold the key to real educational progress in mathematics education.

REFERENCES

Adelman, C. & Alsalam, N.: 1988, December, 'Mathematics study after high school, and its correlates in the careers of young adults', paper presented at Mathematical Sciences Education Board's 'Mathematics Education: Wellspring of U.S. Industrial Strength' Symposium, Irvine, CA.

Airasian, P.W.: 1985, 'Test reviews' [Review of Stanford Achievement Test Forms E and F], *Journal of Educational Measurement*, **22** (2), 163–172.

Alexander, L. & James, H. (Eds.): 1987, *The nation's report card: Improving the assessment of student achievement*, The National Academy of Education, Cambridge, MA.

American College Testing: 1977, *ACT general information*, Iowa City, IA.

Burton, N., Lewis, C. & Robertson, N.: 1988, *Sex differences in SAT scores*, The College Entrance Examination Board, New York.

California Assessment Program: 1989, *A question of thinking: A first look at students' performance on open ended questions in mathematics*, California State Department of Education. Sacramento, CA.

Cameron, R.G.: 1989, *The common yardstick: A case for the SAT*, College Entrance Examination Board, New York.

Cannell, J.J.: 1990, *How public educators cheat on standardized achievement tests*, Friends for Education, Albuquerque, MN.

College Entrance Examination Board: 1988, *1988 National ethnic/sex profiles*, New York.

College Entrance Examination Board: 1989, *Taking the SAT*, New York.

College Entrance Examination Board: 1990, *Major Revisions in SAT approved by trustees. Memo to the members*, New York (October 31, 1990).

Crosswhite, F.J., Dossey, J.A., Swafford, J.O., McKnight, C.C., Cooney, T.J., Downs, F.L., Grouws, D.A. & Weinzweig, A.I.: 1986, *Second international mathematics study: Detailed report for the United States*, Second International Mathematics Study, Urbana, IL.

Crouse, J.: 1986, 'Should a million and a half students be required to take the SAT next year?', *Phi Delta Kappan*, **67** (5), 346–352.

Crouse, J. & Trusheim, D.: 1988, *The case against the SAT*, The University of Chicago Press, Chicago, IL.

Dodge, S.: 1990, 'Average score on verbal section of '89–90 SAT drops to lowest level since 1980; Math score unchanged', *The Cronicle of Higher Education*, September 5, A33–A34.

Dossey, J.A., Mullis, I.V.S., Lindquist, M.M. & Chambers, D.L.: 1988, *The mathematics report card: Are we measuring up?*, Educational Testing Service, Princeton, NJ.

Educational Research Service, Inc. 1981, *Testing for college admissions: Trends and issues*, Arlington, VA.

Freeman, D.J., Kuhs, T.M., Porter, A.C., Floden, R.E., Schmidt, W.H. & Schwille, J.R.: 1983, 'Do textbooks and tests define a national curriculum in elementary school mathematics?', *Elementary School Journal*, **83** (5), 501–513.

Hambleton, R.K.: 1990, *Setting achievement levels for the 1990 NAEP Mathematics Assessment*, NAGB, Washington, DC.

Johnson, W.B. & Parker, A.D.: 1987, *Workforce 2,000: Work and workers for the 21st Century*, Hudson Institute, Indianapolis, IN.

Kulm, G. (Ed.): 1990, *Assessing higher order thinking in mathematics*, American Association for the Advancement of Science, Washington, DC.

Lindquist, M. (Ed.): 1989, *Results from the Fourth Mathematics Assessment of the National Assessment of Educational Progress*, NCTM, Reston, VA.

Mathematical Sciences Education Board: 1989a, *Everybody counts: A report to the nation on the future of mathematics education*, National Academy Press, Washington, DC.

Mathematical Sciences Education Board: 1989b, *Mathematics education: Wellspring of U.S. industrial strength*, Washington, DC.

Mathematical Sciences Education Board: 1990, Grant request for national summit.

McKnight, C.C., Crosswhite, F.J., Dossey, J.A., Kifer, E., Swafford, J.O., Travers, K.J. & Cooney, T.J.: 1987, *The underachieving curriculum: Assessing U.S. school mathematics from an international perspective*, Second International Mathematics Study, Urbana, IL.

Mehrens, W.A. & Lehmann, I.J.: 1987, *Using standardized test in education* (4th ed.), Longman, New York.

Mullis, I.V.S.: 1990, *The NAEP guide*, NAEP, Princeton, NJ.

Mullis, I.V.S., Dossey, J.A., Owen, E.H. & Phillips, G.W.: 1991, *The State Mathematics Achievement: NAEP's 1990 Assessment of the Nation and Trial Assessment of the States*, National Center for Education Statistics, Washington, DC.

Massachusetts Educational Assessment Program: 1989, *On Their Own: Student Responses to Open-Ended Tests in Mathematics*, Massachusetts Department of Education, Quincy, MA.

Massachusetts Educational Assessment Program: 1990, *Beyond Paper & Pencil*, Massachusetts Department of Education, Quincy, MA.

National Assessment of Educational Progress: 1988, *Mathematics Objectives: 1990 Assessment*, Educational Testing Service, Princeton, NJ.

National Commission on Testing and Public Policy: 1990, *From gatekeeper to gateway: Transforming testing in America*, Boston, MA.

National Council of Teachers of Mathematics: 1989, *Curriculum and evaluation standards for school mathematics*, Reston, VA.

Pandey, T. (Ed.): 1991, *A Sampler of Mathematics Assessment*, California Department of Education, Sacramento, CA.

Raizen, S.A. & Jones, L.V. (Eds.): 1985, *Indicators of precollege education in science and mathematics*, National Academy Press, Washington, DC.

Resnick, L.B.: 1987, *Education and learning to think*, National Academy Press, Washington, DC.

Romberg, T.A., Wilson, L. & Khaketla, M.: in press, 'An examination of six standard mathematics tests for grade eight', in Romberg, T.A. (Ed.), *Perspectives on mathematics tests and teaching*, State University of New York Press, Albany, NY.

Romberg, T.A., Zarinnia, A. & Williams, S.: 1989, *The influence of mandated testing on mathematics instruction: Grade 8 teachers' perceptions*, University of Wisconsin-Madison, National Center for Research in Mathematical Sciences Education.

Rosser, P.: 1989, *The SAT gender gap: Identifying the causes*, Center for Women Policy Studies, Washington, DC.

Smyth, F L 1989, 'Commercial coaching and SAT scores. The effects on college preparatory students in private schools', *The Journal of College Admissions*, **123** (Spring, 1989), 2–9.

Wiersma, W & Jurs, S G. 1990, *Educational measurement and testing*, Allyn and Bacon, Needham Heights, MA.

Wilder, G Z & Powell, K 1989, *Sex differences in test performance. A survey of the literature*, The College Entrance Examination Board, New York.

Willingham, W W. & Ramis, L 1982, 'The SAT debate Do Trusheim and Crouse add useful information?', *Phi Delta Kappan*, **64** (3), 207–208.

Wilson, R 1990, 'Students should be coached for admissions tests: True or false?', *The Chronicle of Higher Education*, September 5, A33–A34.

John A. Dossey & Jane O. Swafford
Illinois State University,
Normal, IL
USA

EDWARD A. SILVER & SUZANNE LANE[1]

ASSESSMENT IN THE CONTEXT OF MATHEMATICS INSTRUCTION REFORM: THE DESIGN OF ASSESSMENT IN THE QUASAR PROJECT[2]

1. INTRODUCTION

Recent high-level political interest in the improvement of mathematics education in the United States has led to the increased prominence of reports by the National Academy of Sciences (National Research Council, 1989), the American Association for the Advancement of Science (1989) and National Council of Teachers of Mathematics (1989). These reform-oriented reports have focused the attention of educational practitioners and policy makers on new goals for mathematics education and new descriptions of mathematical proficiency, in which terms like reasoning, communication, problem solving, conceptual understanding, and mathematical power are used frequently to describe an expanded view of mathematical proficiency that goes beyond memorization and mere competence in the basic skills of rational number computation. The reform discussion has thus led naturally to considerations of how to assess students' attainments with respect to this new version of mathematical proficiency and how to assess improvements that may result from curricula and instructional reforms that might be undertaken[3]. This paper focuses on the efforts of one project to deal with the interface between assessment and instructional reform.

QUASAR (*Quantitative Understanding: Amplifying Student Achievement and Reasoning*) is a national project designed to improve mathematics instructional program for students attending middle schools (grades 6–8) in economically disadvantaged communities (Silver, 1989). Currently operating at 6 school sites dispersed across the United States (Silver, 1991), QUASAR seeks to demonstrate that students in these communities can learn a broader range of mathematical content, acquire a deeper and more meaningful understanding of mathematical ideas, and demonstrate an ability to reason and solve appropriately complex problems. When fully implemented, the QUASAR instructional programs will stand in stark contrast to those characterized by what might be called "assembly line" mathematical instruction — the cycle of repetitive drill and practice on basic computation which has characterized middle school mathematics education for many American students. Such instruction has relegated disproportion-

M. Niss (ed.), Cases of Assessment in Mathematics Education, 59–69.

ate numbers of poor students to non-academic programs of study, thereby blocking their access to most socially acceptable paths to status and success[4].

Beyond its goals as practical school demonstration project, QUASAR is also a complex research study of educational change and improvement, in which a major effort will be made to study carefully different approaches to instructional enhancement; to ascertain conditions that appear to be conducive to success; to derive instructional principles for effective mathematics instruction for middle school students; to describe effective instructional programs in ways that will allow their adaptation to other schools; and to devise new assessment tools to measure growth in high-level thinking, reasoning and communication as they relate to mathematics.

Given the goals and aspirations of the QUASAR project, it is imperative that appropriate measures and procedures be developed to monitor and evaluate program impact. One important set of indicators are those that pertain to growth in student knowledge and proficiency over time. Development of the assessments for the QUASAR project has utilized an approach advocated by the National Council of Teachers of Mathematics *Curriculum and Evaluation Standards for School Mathematics* (1989). That report argued for improving the alignment of testing with curriculum goals, advocated the use of multiple sources of assessment information, and suggested that more attention be given both to appropriate methods of assessment and to the proper use of assessment information. With respect to the methods of assessment, the report asserted that an "authentic" assessment of mathematical proficiency would need to address such areas as problem solving, communication, reasoning, and disposition, as well as concepts and procedures.

The QUASAR project uses, or plans to use, a variety of measures in assessing student growth, including paper-and-pencil cognitive assessment tasks administered to individual students in a large-group setting; analysis of student performance on tasks sampling students' individual cognitive activity and their performance in collaborative, small-group settings; analysis of students' performance on tasks, which may involve the use of manipulation materials or computational tools, and which may be relatively brief or which require more extended engagement; and non-cognitive assessments aimed at important attitudes, beliefs, and dispositions.

In the development of assessments, the project has attempted to keep a balanced perspective regarding psychometric constraints and educational needs. This has been possible because the coordinator of assessment development (S. Lane) is a psychometrician by training and the project director (E. Silver) is a mathematics educator. We believe that this balanced perspective is essential for significant progress to be made in establishing alternative assessments as possible replacements for or supplements to the current system of standardized, multiple-choice testing that has become entrenched in the United States. This paper presents an

overview of the design principles for the development of the *QUASAR Cognitive Assessment Instrument* (*QCAI*) — a paper-and-pencil mathematics assessment instrument that is administered by individual students in a large-group setting.

In general, QUASAR assessments are designed to *provide programmatic rather than individual student information*. In other words, we are not attempting to provide indicators for the purpose of assessing individual students; rather, we have designed a system that will collect data from individual students but will provide reliable, valid evaluative information only at the program level. Therefore, the QCAI consists of a relatively large number of assessment tasks (currently about 36) administered at each project site, but each student completes only a small number of tasks (about 9) on each administration occasion. Because of our focus on program evaluation, use of this approach allows us to avoid the difficulty of sampling only a small range of tasks, yet it allows for valid generalization about students' mathematical knowledge and achievement. Over time, it is planned to release some assessment tasks and add new ones. The public release of tasks and scoring rubrics should allow for a clearer understanding of the nature of mathematical proficiencies being assessed and the judgement criteria that are applied in the evaluation of responses. The addition of new tasks each year will allow the QCAI to expand to include not only tasks that reflect important general instructional emphases and topics but also some tasks that may have ben tailored to reflect the unique features of instructional programs that vary across sites. These latter tasks could be developed in close cooperation with the teachers and resource partners at each project site.

Given the goals of the QUASAR project regarding instructional program emphases on breadth of content, QCAI tasks have been developed to assess students' knowledge across a wide range of content areas — extending well beyond whole numbers and arithmetic. Also, given the project's goals related to high-level thinking and deep conceptual understanding, QCAI tasks focus on mathematical reasoning, problem solving, modeling, and communication, and on students' understanding of the features that characterize mathematical concepts and their interrelationships. Due to space limitations the description of the QCAI in this paper will be quite brief in some places. Further details regarding the design principles and conceptual framework for the QCAI can be found in Lane (1991).

2. QUASAR'S ASSESSMENT OF MATHEMATICAL PROFICIENCY: SOME EDUCATIONAL CONSIDERATIONS

The parameters that characterize QUASARS's vision of mathematical ability and mathematical power have been described to a large extent in the *Standards* (National Council of Teachers of Mathematics, 1989), which

suggest the importance of understanding concepts and procedures, becoming a mathematical problem solver, learning to reason mathematically, making connections among mathematical topics and between mathematics and the world outside the mathematics classroom, and learning to communicate mathematical ideas. The vision is also consistent with that of the *Mathematical Sciences Education Board* (National Research Council, 1990) which argued that mathematical power involved the development of the abilities to understand mathematical concepts, principles and procedures, to discern mathematical relations, to reason mathematically, and to apply mathematical concepts, principles, and procedures to solve a variety of non-routine problems.

In this view, mathematics is conceptualized as involving problems that are complex, yield multiple solutions, require judgement and interpretation, require finding structure, and require finding a path for a solution that is not immediately visible. Furthermore, success in mathematical problem solving is viewed as being related to and at least partially dependent on students' beliefs about the nature of mathematics and problem solving, attitudes towards and interest in mathematics, and the socio-cultural context (Lester & Kroll, 1990; Silver, 1985). Specifications for the QCAI assessment tasks were based upon these conceptualizations of mathematical proficiency.

3. QUASAR'S ASSESSMENT OF MATHEMATICAL PROFICIENCY: SOME MEASUREMENT CONSIDERATIONS

An assessment instrument is an imperfect measure of a construct because it either underrepresents the construct domain (i.e., the assessment instrument is too narrow) or in addition to measuring the construct domain it also measures something that is irrelevant to the construct (i.e., irrelevant excess reliable variance), or some combination of the two (Messick, 1989). To ensure that the construct domain is fully represented, QUASAR's assessment of mathematical proficiency is sensitive to many facets, including mathematical *reasoning*, mathematical *communication*, knowledge and use of *strategies* and *representations*, and knowledge and use of mathematical *concepts*, *principles*, and *procedures*. Moreover, the assessment attends to the fact that these interact with various mathematical content areas such as number sense, geometry, and statistics.

Two kinds of construct-irrelevant test variance are proposed by Messick (1989): *construct-irrelevant easiness* and *construct-irrelevant difficulty*. Construct-irrelevant easiness refers to the potential of clues or flaws in the presented task which may allow some students to respond correctly in ways that are irrelevant to the construct domain being measured, and which may lead to scores that are invalidly high. Construct-irrelevant difficulty refers to the possibility that the assessment instrument is, for irrelevant reasons,

more difficult for some groups of students. In designing QUASAR's assessments of students' abilities to think and reason mathematically, we were sensitive to several potential irrelevant constructs that could adversely affect some groups of students, such as differences in reading comprehension ability, writing ability, or familiarity with task contexts. Therefore, the amount and level of reading and writing required of a student was considered in developing the QCAI assessment tasks and scoring rubrics, as was the likely familiarity of the task contexts to the students of differing cultural and ethnic backgrounds. Not only were these two sources of invalidity considered in the process of constructing the assessment tasks and corresponding scoring rubrics but they will also be considered when student performance is interpreted.

Another measurement issue relates to the reliance on a single measure of a complex construct. To triangulate observations of a complex construct, multiple measures are needed. To measure program outcomes and growth in the QUASAR project, the QCAI incorporates a number of task formats (e.g., requiring a student to justify a selected answer versus showing the solution process used to arrive at an answer) and process constraints (e.g., producing a numerical answer versus drawing a diagram). Moreover, as Baker (1990) has noted, any measurement procedure must be understood in the light of other available information and the intended uses of the scores. Therefore, QUASAR also obtains information about classroom instructional processes, students' class assignments and assessments, teachers' knowledge and beliefs about mathematics, and students' beliefs about and disposition towards mathematics. This information can be combined to produce a more complete picture of the performance and attainments of students relative to important program context features.

4. SPECIFICATION OF THE ASSESSMENT TASKS

The development of the QCAI assessment tasks and scoring rubrics involves a collaborative effort by a team consisting of mathematics educators, mathematicians, cognitive psychologists, and psycometricians. Our approach is related to but somewhat different form other examples of alternative assessment frameworks (e.g., Nitko & Lane, 1990; Padey, 1990; Romberg, Zarinnia & Collis, 1990). The QCAI tasks are specified in terms of four components: *cognitive processes*, *mathematical content*, *mode of representation*, and *task context*. With a particular focus on mathematical problem solving and reasoning, the cognitive processes that were specified for task development included the following: *understanding and representing problems*, *discerning mathematical relationships*, *organizing information*, *using procedures*, *strategies and heuristic processes*, *formulating conjectures*, *evaluating the reasonableness of answers*, *generalizing results*, and *justifying answers or procedures*. The content categories included the following:

number and operations (involving decimals, fractions, ratios, proportions); *estimation* (both computational and measurement); *patterns* (both numerical and geometric/spatial patterns); *algebra* (especially tasks related to transition from arithmetic to algebra); *geometry and measurement*; and *data analysis* (including probability and statistics). The range of representations considered in task development include written, pictorial, graphic, tabular, and arithmetic and algebraic symbolic representations. With respect to task context, an attempt was made to embed as many tasks as possible within an appropriate context if this could be done without requiring an excessive amount of reading by the students.

5. SPECIFICATION OF SCORING RUBRICS

A focused *holistic scoring* method is being used to score students' responses to each task. A generalized scoring rubric was designed to incorporate three interrelated components related to the task development specifications described above: mathematical conceptual and procedural knowledge, strategic knowledge, and communication. With respect to *mathematical knowledge*, attention is paid to the extent to which students demonstrate their knowledge of mathematical concepts, principles and procedures, such as understanding relationships among problem elements; using mathematical concepts as a basis for their reasoning, using appropriate mathematical terminology or notation; executing procedures; verifying results of procedures; and generating new procedures or extending familiar procedures. In the area of *strategic knowledge*, attention is paid to students' use of models, diagrams, and symbols to represent and integrate concepts, and their ability to be systematic in applying strategies. The area of *communication* relates to students' ability to convey their mathematical ideas in writing, symbolically, or visually, to use mathematical vocabulary, notation, and structure to represent ideas; to describe mathematical relationships; and to model situations mathematically. Some tasks require the justification of answers; other tasks require the description of strategies or patterns.

The scoring rubrics developed by the *California Assessment Program* (California State Department of Education, 1989) provided a basis for the development of the QCAI generalized rubric. In developing the generalized scoring rubric, criteria representing the three interrelated components were specified for each of five score levels (0–4). Based on the specified criteria at each score level, a specific rubric was developed for each task. The relative emphasis on each component for any specific rubric is dependent upon the particular cognitive demands of the task. In addition to scoring the student responses using the scoring rubric developed for each task, a subset of the student responses will be carefully analyzed to provide more detailed information regarding the types of representation and strategies students use, the nature of errors or misconceptions in students' work, and

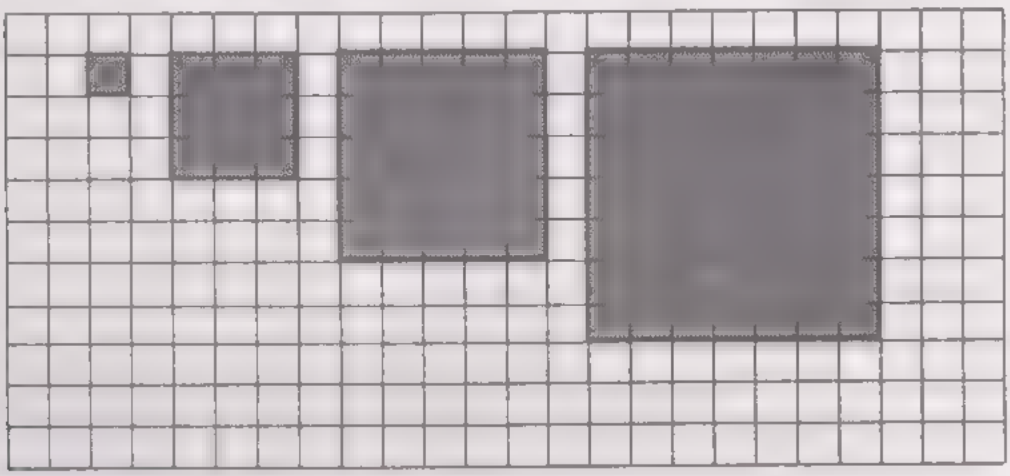
the nature of the mathematical knowledge and cognitive processes underlying successful performance.

6. SAMPLE TASKS AND ADMINISTRATION INFORMATION

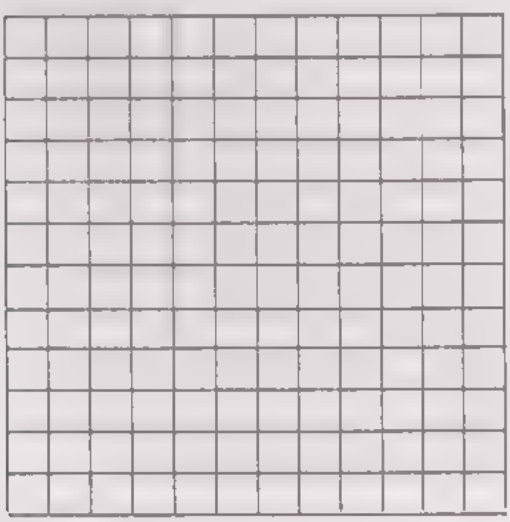
For the 1990–1991 school year, a set of thirty-six assessment tasks was developed for use with 6th Grade students. The set of thirty-six tasks was divided into four sets of nine different tasks, which were randomly distributed to students in each classroom. Students received a different set in each of the Fall and Spring administrations. Two examples of assessment tasks similar to those used in QCAI are provided in Figure 1.

Task 1 - Mathematical Content: Pattern recognition

Look at the following pattern of figures:



A Draw the 5th figure



B. Describe the pattern.

Task 2 - Mathematical Content: Numbers and Operations
The table below shows the cost for different bus fares.

BUSY BUS COMPANY FARES	
One Way	$ 1.00
Weekly Pass	$ 9.00

Yvonne is trying to decide whether she should buy a weekly bus pass On Monday, Wednesday and Friday she rides the bus to and from work On Tuesday and Thursday she rides the bus to work, but gets a ride home with her friends.

Should Yvonne buy a weekly bus pass? ____________

Explain your answer.

Figure 1 *Sample assessment tasks*

For the first task, it is expected that a student would draw a 9-by-9 square on the grid provided and shade the square in. Also it is expected that a student would describe the pattern by saying "It is a pattern of squares with odd sides — 1,3,5,7,9,11, and so on'; or "In the pattern you add 2 rows and 2 columns to each square to get the next square"; or some other similar description. In the next task, we would expect that a student's response would show evidence of a clear reasoning process. For example, a student might answer "no" and provide an explanation, such as "Yvonne takes the bus eight times in the week, and this would cost $8.00. Since the bus pass costs $9.00, she should not buy the pass". It is possible, however, that a student might answer "yes" and provide a logical reason such as "Yvonne should buy the bus pass because she rides the bus eight times and this costs $8.00. If she rides the bus on weekends (to go shopping, etc.), it would cost $2.00 or more, and that would be more than $9.00 altogether, so she can save money with the bus pass". As this example suggests, task presented in this open-ended format may allow for more than one possible correct answer.

A task actually contained in the QCAI assessment for the 1990–91 school year is the following:

Yolanda was telling her brother Damian about what she did in math class. Yolanda said, "Damian, I used blocks in math class today. When I grouped the blocks in groups of 2, I had one block left over. When I grouped the blocks in groups of 3, I had 1 block left over. And when I grouped the blocks in groups of 4, I still had 1 block left over."

Damian asked, "how many blocks did you have?"

What was Yolanda's answer to her brother's question? Show your work.

Figure 2 *Sample assessment task*

On this problem, it is expected that students will produce an answer that simultaneously satisfies all the problem constraints, and that they will provide information about their solution method (e.g., systematic guess and test). For this problem, multiple answers are possible, and multiple solution methods can be utilized by the students.

After student responses have been obtained, the papers are scored by teams of classroom teachers who are trained as raters. The raters use the scoring rubric for each task in order to assign a score between 0 and 4 to each student's response. In addition to these holistic judgements, student responses for a sample of the tasks are subjected to further examination and analysis in order to identify cognitive process information, data regarding strategy usage, systematic error patterns, and other important insights related to the mathematical knowledge and performance of the students. The general performance information for all tasks and the detailed reports on selected tasks will be summarized in reports that can be shared with the teachers at the project sites.

As noted earlier, QUASAR intends to use a wide range of assessment procedures. For example, QUASAR is supplementing the cognitive information obtained from the QCAI with non-cognitive assessments aimed at important student attitudes, beliefs, and dispositions. With respect to cognitive measures, in addition to the group-administered QCAI tasks which measure students' individual cognitive activity, QUASAR will also attempt to analyze students' performance in collaborative, small-group settings, since cooperative learning and small-group problem solving are instructional practices used frequently by QUASAR teachers. Beyond using the QCAI tasks, which measure students' performance on paper-and-pencil tasks completed during a relatively short time span (about 5 minutes for each task), QUASAR will also try to analyze students' performance on tasks which involve the use of manipulative materials or computational tools (e.g., calculators), and on tasks which may require intellectual engagement over a more extended period of time. It is hoped that samples of student work (e.g., tests, classwork, homework, projects) supplied by teachers at the project sites will provide the data for these supplemental analyses, and that these instructionally-embedded assessment data will

provide another indicator of the nature and extent of intellectual activity in the classrooms and supplement the information obtained from the QCAI regarding students' developing mathematical proficiencies.

NOTE

1. Each author contributed equally to the conceptualization of this paper. In fact, the design of the QCAI discussed herein is largely the product of Lane's intellectual leadership of QUASAR's assessment work over several years. Because Silver received the invitation to participate in the ICMI conference, he is listed as first author.

2. Preparation of this paper was supported by a grant from the Ford Foundation (grant number 890–0572) for the QUASAR project. Any opinions expressed herein are those of the authors and do not necessarily reflect the views of the Ford Foundation.

3. An expanded discussion of the role of assessment in the mathematics education reform movement in the United States, including the prevalence and limitations of an assessment-driven reform strategy, can be found in Silver (in press).

4. The substantial neglect of high-quality education for poor children in the United States has been well documented. Although there are some exceptional examples of schools that have done an effective job of providing a thoughtful education to the children of poverty, the general finding is that schools in poor communities are characterized by impoverished resources, poor organizational structure and quality, an underprepared teaching staff, classroom instruction that focuses almost exclusively on low-level knowledge and skills, and abysmal student achievement (Kozol, 1991; Natriello, McDill & Pallas, 1990). Although QUASAR faces many obstacles to substantial instructional reform in these schools, the project posits that the prior failures can be overcome with effort, imagination and the judicious application of modest financial resources.

REFERENCES

American Association for the Advancement of Science: 1989, *Project 2061: Science for all Americans*, Washington, DC.

Baker, E.L.: 1990, 'Developing comprehensive assessment of higher order thinking', in Kulm, G. (Ed.), *Assessing higher order thinking in mathematics*, 7–20, American Association for the Adcancement of Science, Washington, DC.

California State Department of Education: 1989, *A question of thinking: A first look at students' performance on open ended questions in mathematics*, Sacramento, CA.

Kozol, J.: 1991, *Savage inequalities*, Crown Publishers, New York.

Lane, S.: 1991, April, *The conceptual framework for the developments of a mathematics assessment instrument for QUASAR*, paper presented at annual meeting of the American Educational Research Association, Chicago, IL.

Lester, F.K., Jr. & Kroll, D.L.: 1990, 'Assessing student growth in mathematical problem solving', in Kulm, G. (Ed.), *Assessing higher order thinking in mathematics*, 53–79, American Association for the Advancement of Science, Washington, DC.

Mathematical Sciences Education Board: 1990, *Reshaping school mathematics: A philosophy and framework for curriculum* National Academy of Sciences, Washington, DC.

Messick, S.: 1989, 'Test Validity', in Linn, R.L. (Ed.), *Educational measurement* (3rd ed.), 13–104, American Council on Education, New York.

National Council of Teachers of Mathematics 1989, *Curriculum and evaluation standards for school mathematics*, NCTM, Reston, VA.

National Research Council. 1989, *Everybody counts*, National Academy of Sciences, Washington, DC.

Natriello, G., McDill, E.L. & Pallas, A.M. 1990, *Schooling disadvantaged children. Racing against catastrophe*, Teachers College Press, New York.

Nitkon, A.J. & Lane, S. 1990, August, *Solving problems is not enough. Assessing and diagnosing the ways in which students organize*, paper presented at the Third International Conference on Teaching Statistics, Dunedin, New Zealand.

Pandey, T. 1990. 'Power items and the alignment of curriculum and assessment', in Kulm, G. (Ed.) *Assessing higher order of thinking in mathematics*, 39–52, American Association for the Advancement of Science, Washington, DC.

Romberg, T.A., Zarinnia, E.A. & Collis, K.F. 1990, 'A new world view of assessment in mathematics', in Kulm, G. (Ed.), *Assessing higher order thinking in mathematics*, 21–38, American Association for the Advancement of Science, Washington, DC.

Silver, E.A. in press, 'Assessment and mathematics education reform in the United States', to appear in *International Journal of Educational Research*.

Silver, E.A. 1985, 'Research on teaching mathematical problem solving: Some underrepresented themes and needed directions', in Silver, E.A. (Ed.), *Teaching and learning mathematical problem solving: Multiple research perspectives*, 247–266, Lawrence Erlbaum Associates, Hillsdale, NJ.

Silver, E.A. 1989, QUASAR *The Ford Foundation Letter*, 20(3), 1–3

Silver, E.A.: 1991, *QUASAR (Quantitative Understanding: Amplifying Student Achievement and Understanding) project summary*, Learning Research and Development Center, University of Pittsburgh.

Edward A. Silver & Suzanne Lane
Learning Research and Development Center,
University of Pittsburgh,
USA

MARGARET BROWN

ASSESSMENT IN MATHEMATICS EDUCATION: DEVELOPMENTS IN PHILOSOPHY AND PRACTICE IN THE UNITED KINGDOM

1. INTRODUCTION

The last ten years have seen a remarkable change in the nature of assessment in the UK. While this has occurred in all subjects, mathematics has been at the forefront of policy formation, research and development.

The *Cockcroft Report* (DES, 1982) initiated broader modes of summative assessment in our examinations at age 16, introducing extended problem-solving tasks assessed by teachers alongside the written examination papers. The second section of this paper will report briefly on these changes.

Another strand of development comes in relation to continuous diagnostic assessment carried out by both primary (ages 5-11) and secondary (ages 11-16) teachers on a criterion-referenced model, but using a basis of cognitively based strategies rather than technical skills. This will form the subject of the third section.

The most recent and most profound change has been the introduction of a *national assessment system*, reporting publicly at ages 7,11,14 and 16, within a single framework of 10 criterion-referenced levels common to all ages. This initially both brought together and standardized the summative and diagnostic aspects discussed above. It introduced also nationally standardized extended tasks to assess simultaneously both process and content over a wide range of attainment. These developments, together with the shifts in government policy which are continuing to steer them, form the focus of the fourth section.

2. WIDENING OF ASSESSMENT MODES IN SUMMATIVE ASSESSMENT AT AGE 16

In the mid-seventies, the Government, concerned about standards of achievement, instituted the *Assessment of Performance Unit* (APU) to undertake national monitoring at ages 11 and 15. In addition to written items, this incorporated an element of practical testing of concepts and skills in an oral mode, using one-to-one interviews. Examples included estimating and measuring the lengths of some curved lines, tests of

M. Niss (ed.), Cases of Assessment in Mathematics Education, 71–84.

calculator skills, and probability experiments. To start with, tests were "practical" in the sense of using equipment, but in 1980 they began to broaden out to include problem solving skills, mainly in real contexts - for example, using timetables and maps to plan a day-trip (Foxman, 1987).

The rationale for such practical work lay mainly in the move towards a utilitarian curriculum which took place in the nineteen-seventies, following what was perceived as the purer excesses of "modern mathematics" in the sixties. Practical tests were later incorporated in some local systems of testing for 11 year olds and in local and national tests for low attaining students at 16 who were not catered for in the system of public examinations which then existed (e.g. SMP (the *School Mathematics' Project*) Graduated Assessment, see Close & Brown, 1988, 1990).

The Cockcroft Report, in 1982, institutionalized practical mathematics and problem solving as two of the six required modes of experience which should form part of mathematics teaching. Pure mathematical investigation was also included among the six, a surviving minority development from the sixties which had been carefully nurtured by the Association of Teachers of Mathematics (ATM).

As part of an integrated set of recommendations covering many different aspects of mathematics teaching, the Report stated:

> 'Examinations in mathematics which consist only of timed written papers cannot, by their nature, assess ability to undertake practical and investigational work or ability to carry out work of an extended nature. They cannot assess skills of mental computation or ability to discuss mathematics nor, other than in very limited ways, qualities of perseverance and inventiveness. Work and qualities of this kind can only be assessed in the classroom and such assessment needs to be made over an extended period." (DES, 1982, paragraph 532)

The result was that when our public examinations at 16 were reformed to produce one system catering for most children, known as the *GCSE* (*General Certificate of Secondary Education*), the *National Criteria* for examinations in mathematics required the inclusion of oral and practical work, and of teacher-assessed extended projects of an investigational nature. Following the appointment of Sir Wilfred Cockcroft as the Chairman and Chief Executive of the new *Secondary Examinations Council*, this type of requirement was made in all subjects.

The new GCSE was first examined in 1988, with the new elements becoming compulsory only in Summer 1991. However most schools were setting and marking extended work by 1989 or early 1990 since four or five tasks are usually spread out over a two-year interval.

Our public examinations at 16 are conducted by five *Examination Groups*, which are independent commercial enterprises competing for custom among schools. The Examination Groups work within a framework of National Criteria approved by the Government which has allowed them to differ in the degree of structure they have provided for extended work.

Some Groups gave teachers complete freedom over the topics and marking schemes, relying on moderators to adjust standards where necessary.

Some specified broad topics (e.g. one project of a "real life" nature involving geometrical design, a number investigation, a statistical survey) and have given broad indications of what is required for each grade (in terms such as "collecting and representing results systematically", "finding a number pattern in a table of results", "expressing a generalization algebraically", and so on).

Other Examination Groups have presented much more tightly structured tasks with specific marking schemes. These take the form of both pure mathematical investigations which require results to be generated and generalizations found, and practical problems using resources which are provided or of a standard kind (e.g. store catalogues, product packaging, etc.).

An important part of the implementation of the Cockcroft Report was the appointment of advisory teachers in each local education authority to work in groups of schools helping teachers to introduce and assess practical and investigational work.

Many good sets of teaching and assessment materials have been produced, among the most popular being those published by the *Shell Centre*/Joint Matriculation Board (1984-90), *SMP* (1989), *West Sussex Institute of Higher Education* (Ahmed & Bufton, 1986), and *GAIM* (*Graded Assessment in Mathematics Project*) (GAIM, 1988; Brown, in press).

Mental and oral work have also been introduced into the examination system, using class tests, or, less frequently, interviews. Some of the extended work regulations allow, or even encourage, groupwork; this may also provide a forum for assessing pupils' contributions to discussion.

The result has been curriculum development and implementation on a national scale, broadening the style of classroom work to encompass extended investigative work, 'real' problem-solving, oral and practical activity and groupwork. These developments are now spreading to courses and matching examinations for the 16-19 age group (e.g. Dolan, 1991). While it is difficult to isolate the importance of the various recommendations of the Cockcroft Report, the fact that primary schools have lagged well behind secondary and that many schools have delayed making changes until it was an examination requirement, suggest that to a considerable extent the change has been assessment-led.

Of course the quality of teaching is extremely variable, with some teachers "teaching rules for doing investigations" in the same mindless way in which they taught routine algorithms for subtraction. Nevertheless much of the response has been very positive; Her Majesty's Inspectors see the broadening of GCSE styles of examining as a major factor in improving the quality of mathematics teaching (DES, 1991).

Nevertheless, a new Prime Minister and Secretary of State have expressed doubts about the "objectivity" of such coursework assessment, and have decreed against all professional advice that it can from 1994 be used for not more than 20 percent of the assessment at 16.

3. CRITERION-REFERENCED COGNITIVELY-BASED DIAGNOSTIC ASSESSMENT

Criterion-referencing took some time to cross the Atlantic. It only eventually did so by being transformed into a model that was based, not on the narrow learning of technical skills, but on cognitive strategies, in areas of content and process, more closely in tune with the prevailing broadly constructivist philosophies of mathematics education in Britain.

It arrived first in Scotland, where it was introduced into the reformed *Standard Grade examinations* at 16+, in particular into the teacher-assessed component. Some useful research occurred (e g. Black & Dockrell, 1984) but not much in mathematics. However, the scope of the initiative had to be greatly reduced as teachers found it unmanageable.

The notion of assessment as providing a description of a child's attainment, rather than only a grade in comparison to some age-group norm, then travelled south, having influenced the then Secretary of State for Education in England. He asked for 'Grade Criteria' to be provided for the new GCSE at 16, largely so that parents and employers would gain useful information and so that national standards of performance could better be monitored (DES, 1985). Attempts were made to do this (Secondary Examinations Council, 1985) but foundered largely because of the problem of assessing a broad and extensive set of criteria in an examination which, in spite of the changes described in the last section, was still dominated by performance in short written examinations.

Nevertheless the Secretary of State's intention to extend criterion-referencing to primary level made clear in the same document (DES, 1985) was to form the basis for the *National Curriculum Assessment*.

Although criterion-referencing gained little ground in formal examinations, the notion of describing attainment ("profiling") was taken up by another movement in England and Wales. This aimed to provide each secondary school student with an ongoing *Record of Achievement*, containing a portfolio of "best work" and positive descriptions of all aspects of achievement, including personal and social skills and extra-curricular activities. This was to be the subject of ongoing negotiation with pupils and parents, and to result in a summarized document to be taken away by school leavers. Records of Achievement were backed by the Government, although without great enthusiasm, and will in a diluted form be compulsory for secondary school leavers and introduced into primary schools from 1991.

Although mathematics is a strand in the pilot Records of Achievement development schemes in several local education authorities, the only one to have lasting national influence is probably that in London.

This scheme, *Graded Assessment in Mathematics* (*GAIM*), was one of five covering the major curriculum areas, and was based on the notion of progressive levels of attainment, pioneered successfully in modern languages. It was a partnership of three agencies, the Inner London Education Authority (ILEA), the University of London Examination and Assessment Council (one of the Examinations groups for GCSE), and King's College London.

The College brought experience of cognitively based diagnostic assessment as part of the *Chelsea CSMS* (*Concepts in Secondary Mathematics and Science*) project (Hart et al., 1981, 1984). The Examination Group backed the scheme to qualify as an alternative route to GCSE, using continuous teacher assessment with visiting moderators instead of externally set and marked terminal timed written tests. The ILEA had been the first to develop a diagnostic assessment scheme which operated at primary level (Checkpoints) (ILEA, 1976).

GAIM has a framework of criteria organised into progressive levels of attainment. Although some of these criteria are fairly routine in nature to satisfy the GCSE National Criteria, others are related either to logical thinking or to content-based cognitive strategies identified by earlier research (Brown, 1989; in press). For example two criteria in different areas and at different levels of attainment are:

- Can take into account two constraints or attributes when classifying, planning, inventing or problem-solving, and can check results.

- Can use multiplication and division, on a calculator if necessary, to solve problems involving rates using numbers of any size.

(Each criterion is accompanied by several examples to make its meaning clearer.)

Other schemes share aspects of GAIM (e.g. the *Oxford Certificate of Educational Achievement*, the *SMP Graduated Assessment* scheme, the Shell Centre/JMB *Numeracy through Problem Solving*, the *Association of Teachers of Mathematics* GCSE scheme). Nevertheless GAIM is probably the most radical in terms of encouraging teachers to implement a rigorous criterion-referenced assessment system using a variety of means of assessment. Recommended assessment methods emphasize open activities developed and produced by the project which are of a problem-solving and investigatory kind (both pure and applied) and integrate the use of criteria. They also include teacher-verified student self-assessment, and observation and discussion, as well as the more usual classroom tasks and tests.

Teachers have found operating such a system over the whole 11-16 age group to be a complex but rewarding task. In GAIM and in similar schemes there have been considerable gains in a number of areas (Close & Brown, op.cit). Teacher professionalism has increased, with teachers becoming much more aware both of the nature of the mathematics they are teaching and of their students' individual achievements and weaknesses. This has often resulted in provision of a curriculum which is more appropriate to students' needs.

While a need has been demonstrated in most of these schemes for teachers from different schools to meet regularly to agree on shared meanings and to converge in their practices (Love & Shiu, 1991), this has been a powerful agent in accelerating professional development.

Since students have become more involved in their own assessment and learning, with their own versions of the criteria so that they can see what has to be achieved to obtain the next level, student motivation has sometimes risen dramatically. Schools have correspondingly found that their GCSE results have improved.

The recent Government decision, referred to at the end of the previous section, that written examinations, externally set and marked, must again account for at least 80 percent of marks in the GCSE, is likely to reduce the motivation of students and teachers, and to increase the drop-out rate.

4. NATIONAL ASSESSMENT: COMBINING FORMATIVE ASSESSMENT WITH NATIONALLY STANDARDIZED TESTING

In 1987 the British Government made the decision to introduce a *National Curriculum*. This would be legally binding, with national testing of all pupils at the end of each of four *Key Stages* (ages 7, 11, 14 and 16), and public reporting of the results of each school and local authority in the form of a league table. The curriculum was to be operating for mathematics and science in all schools by 1989, with the first national testing of 7-year-olds in 1991, of 14-year-olds in 1992 (now postponed until 1993), and of 11-year-olds in 1994.

The original notion had been a set of attainment targets in each subject for each of these four key stages; the Government-appointed *Task Group on Assessment and Testing* (TGAT) (1988) succeeded in changing the model, adopting that used by graded assessment in which a series of progressive age-independent levels is defined, with the assessments reporting each pupil's progress through these fixed *attainment levels*.

There are 10 levels defined for the National Curriculum, each representing a notional two years' progress for an average student, with level 1 defined as representing roughly the attainment of an average 5-year-old, level 6 that of an average 15-year-old, and level 10 that of the top 2 or 3 percent of 16-year-olds.

Each subject is divided into a number of strands called *Attainment Targets*; each attainment target is then defined by *statements of attainment* (criteria) at each level. For example, after a recent modification, mathematics now has five attainment targets, one for *process* (*Using and Applying Mathematics*), and four for *content* (*Number, Algebra, Shape and Space*, and *Handling Data*).

Each attainment target has on average three statements of attainment to define each of the ten levels of attainment, but occasionally there are as many as five. The statements are rather more general than those illustrated earlier from the GAIM project; for example:

"Find ways of overcoming difficulties when solving problems." (Using and Applying, level 3)

"Solve numerical problems, checking that the results are of the right order of magnitude." (Number, level 8)

The results of national assessment will be reported in the form of subject profiles; each student will receive a level between 1-10 for each attainment target in each subject.

The TGAT group also initially steered the Government away from the original intention of written short-item tests in each subject at the end of each key stage, and towards a model which incorporated both of the two aspects of assessment discussed above.

The TGAT group proposed that continuous teacher assessment would be the basis for the students' results, with *Standard Assessment Tasks* (*SATs*), taken at the end of each key stage used only to moderate teachers' judgements (in the sense that if a wide discrepancy occurred, the teacher--assessed levels would be re-examined and if they could not be substantiated, would be altered).

Building on the GCSE experience reported above, the standard assessment tasks (SATs) were recommended by TGAT to contain a broad spread of testing modes, including extended projects, oral and practical work, and, in primary schools, crosscurricular theme-based tasks.

There are still political battles being fought about the two issues above, i.e.

- what is the relative emphasis to be placed on continuous teacher assessment and on terminal SAT results in arriving at the reported results?
- what range of assessment modes is appropriate for SATs?

For example, in 1991 the first national round of testing took place at *key stage 1 (age 7)*. The arrangements were that:

- all teachers of 7-year-olds submitted by March 31st their own assessments of each child on each attainment target of English, science and mathematics; each attainment target was assessed as a level in the range 0-4, with each level in each attainment target defined by the associated statements of attainment.
- over a 3-week period in May each child was tested, using SATs which had previously been distributed to schools. In mathematics this involved three parts, one on number operations, one on using and applying mathematics in "real" problem-solving and practical investigations, and the choice for the third one included shape properties and data handling.
- the teacher assessed results were generally replaced by the SAT test results in each of the attainment targets in which SAT results were available. Appeals that the SAT results were invalid, and the teacher assessed results should be taken instead, were possible, but rarely made.

As part of this national assessment, each teacher with 7-year-olds in her class, and in most cases each headteacher, had three days of training for carrying out teacher assessment, and for administering and marking the SATs.

Evaluation reports (e.g. DES, 1991; Gipps et al., 1991) indicate that teachers were as a result being much more systematic throughout the year about their curriculum planning and assessment. Although the notion of assessing the individual strengths and weaknesses of children was familiar to teachers of 7-year-olds, teachers said that the presence of criteria statements helped them to focus their assessment better. Nevertheless they felt the statements were sometimes too vague. This coincides with reports of another study (Frobisher & Nelson, 1991) which suggest, not surprisingly, that different ways of interpreting the statements can produce very different results.

The key stage 1 (age 7) SATs, developed by the *National Foundation for Educational Research* (*NFER*), were expected to be treated like any other classroom activities. Most were carried out in an oral mode with a small group of children. Teachers had some flexibility in how they were administered and whether they were adapted to fit in with a particular theme going on in the classroom.

A few SAT activities, such as the group task for the attainment target in Using and Applying Mathematics of inventing a game which required addition or subtraction of dice-scores, were given to most or all children and assessed by outcome. Most activities however were tied to particular criteria statements at levels 1,2 or 3 and the teacher had to make decisions on the basis of her teacher assessment as to which level of activities to give to each child. The child was then tested at the next higher or lower level depending on whether the child was successful at the entry level.

Those reports which are publicly available (e.g. Gipps et al., op cit) demonstrate that teachers were generally happy about the quality of the SATs as classroom activities. Some showed that they had learned useful things about their pupils' attainments (SAT results differed from teachers' assessments in a third of cases). Some felt that their classroom practice had been enhanced by using open-ended tasks in mathematics and science for the first time.

However, all teachers found the organization, assessment and recording to be formidable, and were concerned at the lack of attention given to the children who were not at that moment being assessed.. In Scotland, where the tests were then not a legal requirement as they were in England and Wales, large numbers of parents refused to allow their children to take them.

The new Secretary of State for Education has recently replaced the chair of the *School Examination and Assessment Council*, and has used the problems teachers expressed over classroom organisation in 1991 as a reason for introducing modifications to the procedures for 1992. Almost all the mathematics SATs are now in the form of worksheets with routine pencil-and-paper items (often straightforward "sums" without any everyday context), which can be done by the whole class at once. The process attainment target (Using and Applying Mathematics) will no longer be assessed by external task, thus removing one of the components most effective in encouraging professional development in 1991.

The response from teachers is that the organisation of the new SATs will be little easier since they still prefer to administer them to small groups. Educationally they feel that the new tasks are backward-looking and encourage rote learning and 'teaching to the test". In addition, short term coaching is likely to be encouraged by a new law which requires the publication of schools results in league tables.

At the end of *key stage 3 (age 14)*, where the first full national assessment has been postponed until summer 1993, the developments have been similar, but with more extreme shifts of policy .

The contract for developing the SATs in mathematics is held by King's College London with a team directed by Gill Close, many of whom previously worked on the Graded Assessment in Mathematics (GAIM) project.

At age 14, the form of the assessment piloted in 1990 and 1991 was more experimental, with open tasks, either "real world" problems (e.g. running a food stall at a school fund-raising event) or mathematical investigations (e.g. investigating patterns in shapes made from closed loops of octagons), being used to assess both content and process. Each task was intended to occupy mathematics lessons over a 2-3 week period, and to assess one process and about two content attainment targets, covering criteria statements in each of the levels 1-10. (Level 1, although defined at an average 5-year-old level, is appropriate for some 14-year-olds with

severe learning difficulties; equally not more than 1 percent of the population would be expected to have reached level 10). An optional computer-based version which was made available was found particularly helpful by students with severe motor difficulties.

Extended open activities were chosen so that students at different attainment levels could be engaged on the same task; the alternative of providing tasks at each level in each attainment target to be administered individually to different students was felt to be administratively too daunting.

Students were encouraged to tackle the task using as powerful mathematical ideas and skills as they could. Part way through the task they were provided with a "personal target check" in the form of a list of the criteria statements at the relevant levels in the assessed attainment targets, expressed in appropriate language. This was to enable students to check that they had demonstrated as many of these statements as they could in the task. Thus the student self-assessment being encouraged as part of the formative teacher assessment (SEAC, 1991) was being extended to the SAT assessment.

In addition a few of the relevant criteria statements which students were unlikely to demonstrate spontaneously in the open task were tested in focused written or oral items related to the theme of the activity (e.g. calculating dimensions of the octagons which are not directly needed in the tiling investigation). Pupils completed only the items on relevant levels.

Although part of the initial activity and discussion of the task took place in groups, each student wrote their own report. This, together with the less tangible behavior in the classroom, was assessed by the teacher, in the form of a level for each attainment target assessed, according to a marking scheme which indicated how the statements of attainment could be demonstrated in that particular SAT activity.

The 1991 pilot involved 20,000 pupils in 161 schools, including 13 schools for pupils with special educational needs of one kind or another. The tasks were generally well-received by teachers, and had an enthusiastic reception both among students and in the mathematics education community. Eighty-seven percent of pupils said they had enjoyed the work, and 90 percent felt that they had learned some mathematics. This was borne out by their teachers and by observers. For each task more than three times more pupils wanted to continue to work on the project for longer than three weeks as thought the time involved was too long.

Teachers reported that the SAT took no more preparation than normal classwork. Although the assessment took longer than normal class assessment for this period (27 minutes per pupil against 21), it took less long than GCSE coursework assessment (at 29 minutes). This was in spite of the fact that 42 percent of the teachers had had no previous experience of assessing against criteria. The hardest part of the assessment for teachers was the need to assess some of the oral and practical aspects during class

time, which only 49 percent found to be manageable; nevertheless, where it was possible it was generally agreed to be both more accurate and more beneficial than written assessment.

Shortly before the pilot studies were undertaken, the new Secretary of State, being concerned about the complaints at teacher workload at key stage 1 (age 7), examined the pilot assessment materials for mathematics and science at key stage 3 and declared to the national press that they were "elaborate nonsense". Without waiting for the results of the pilot, he announced that the national tests for 14-year-olds in 1992 would be in the form of short written tests, and modified the development contracts for all subjects at key stage 3.

The current position is that 75 percent of schools are expected to take part in the 1992 round of testing. (Participation cannot, as intended, be legally required until 1993 since the modified attainment targets on which the tests will be set cannot legally come into force before September 1992). Each pupil will sit three one hour tests on June 9th at specified times, and between them the tests will provide a level on all four of the content attainment targets. At least half the statements of attainment will be assessed at each level on a criterion-referenced basis, with the marking being carried out by teachers and audited by examination board personnel.

The items are allocated into four bands of adjacent levels: 1-4, 3-6, 5-8 and 7-10. Pupils are entered for the band of levels which teachers judge is most appropriate on the basis of their teacher assessments, and will be assessed at a level in the agreed range for each attainment target. Apart from the entry decision, it seems likely that teacher assessment results for the content targets will be used only in case of appeal.

The process attainment target, Using and Applying Mathematics, will be assessed only by teacher assessment. A set of the materials used for the pilot in 1991 will probably be sent to all teachers to assist this assessment, but teachers may choose to use other methods of assessment.

One of the advantages of the SATs piloted in 1991 was that the content areas were assessed as part of an extended task. This meant that pupils had to be genuinely able to apply their understanding and skills in a problem--solving context. However creative the team, the return to written tests with short items, as at key stage 1, makes it easier to coach pupils superficially for the tests, leading back to the barren curriculum which has been the result of examination-oriented mathematics teaching in the past.

Due to these changes, teachers at secondary level have not all yet started on the continuous criterion-referenced teacher assessment. This has additional complications at this level, partly because teachers do not see their students so frequently as at primary level (although primary teachers have more subjects to cover in return). They are also in the position over the next few years of having to determine on which levels to place 11-year-old students in each attainment target; by 1994 students will arrive at secondary schools with a comprehensive record of previous attainment.

One of the problems already found at primary level which seems likely to recur in a more extreme form at secondary level is that of teachers finding it difficult to distinguish curriculum coverage from permanent learning ("the implemented curriculum" from "the attained curriculum", e.g. Travers, 1989). At age 7, results of 1991's national tests indicate that teachers tended to over-estimate their teacher-assessed results as compared with the SATs results in the much-taught areas of number and measurement, and correspondingly to underestimate in less-taught areas of shape and data-handling.

5. CONCLUSIONS

The assessment of the national curriculum initially brought together, using a new level-based criterion-referenced framework, the two previous trends in assessment in Britain:

- broadening of modes of summative assessment,
- continuous formative/diagnostic assessment by teachers.

Each of these changes were themselves novel and neither universally implemented nor fully evaluated. Due to the undue haste of the Government we thus embarked on a huge national assessment experiment with insufficient consultation, planning or trialling. Nevertheless initial indications suggested that at least some aspects were successful and brought genuine educational advantage. Teachers favoured modifications leading to a system that an earlier feasibility study had proposed (Denvir et al., 1987). This was that they should be left free to give SATs at any time, and to take them into account in arriving at their own assessments, which would be reported, subject to moderation.

But before there had been time to evaluate the system, yet another series of irrational decisions emanating from new ministers in the same Government have reversed much of the previous policy. The UK now seems to be heading back to our previous position where the curriculum becomes subservient to the requirements of regular routine written examinations, which the Cockcroft Committee in 1982 identified as a major cause of low standards of motivation and achievement. We have on board a disillusioned set of teachers and educationists who put a great deal of now apparently wasted effort into realizing the more positive effects of the earlier proposals.

Other countries can hopefully learn from our failure to take our politicians with us on the journey.

REFERENCES

Ahmed, A.G. & Bufton, N.: 1986, *Teachers Evaluating and Assessing Mathematics*, Southern Regional Examinations Board/West Sussex Institute of Higher Education.

Black, H.D. & Dockrell, W.B.: 1984, *Criterion-referenced Assessment in the Classroom*, Scottish Council for Research in Education, Edinburgh.

Brown, M.: 1989, 'The Graded Assessment in Mathematics project', in Robitaille, D.F. (Ed.) *Evaluation and Assessment in Mathematics*, Science and Technology Education Series No. 32, UNESCO, Paris.

Brown, M. (Ed.): in press, *Graded Assessment in Mathematics Pack*, Macmillan Education.

Close, G. & Brown, M.: 1988, 1989, *Graduated Assessment in Mathematics. Report of the SSCC Study. Parts 1 & 2*, Department of Education and Science, London.

Denvir, B., Brown, M. & Eve, P. 1987, *Attainment Targets and Assessment in the Primary Phase. Mathematics Feasibility Study*, Department of Education and Science.

Department of Education and Science (DES)/Welsh Office, Committee of Inquiry into the Teaching of Mathematics in Schools. 1982, *Mathematics Counts* ('The Cockcroft Report'), Her Majesty's Stationery Office, London.

Department of Education and Science (DES)/Welsh Office: 1985, *Better Schools (CMND 9469)*, Her Majesty's Stationery Office, London.

Department of Education and Science (DES)/Welsh Office, Task Group on Assessment and Testing (TGAT). 1988, *A Report*, Her Majesty's Stationery Office, London.

Department of Education and Science (DES)/Her Majesty's Inspectorate: 1991, *Mathematics Key Stages 1 and 3, a report by HM Inspectorate on the first year 1989-90*, Her Majesty's Stationery Office, London.

Dolan, S.: 1991, 'Assessment in 16-19 mathematics', in Pre-proceedings of ICME Conference Assessment in Mathematics Education and its Effects, Calonge, Spain, 1991

Hart, K. (Ed.): 1981, *Children's Understanding of Mathematics: 11 16*, John Murray

Hart, K., Kuchemann, D.E. & Brown, M.: 1984, *Chelsea Diagnostic Tests*, NFER-Nelson, Windsor, UK.

Foxman, D.: 1987, *Assessing Practical Mathematics in Secondary Schools*: Assessment of Performance Unit Series, Department of Education and Science. NFER-Nelson, Windsor, UK.

Frobisher, L. & Nelson, N. 1991, 'An evaluation of the national curriculum assessment in mathematics of seven year old children in England and Wales', in Pre-proceedings of ICME Conference Assessment in Mathematics Education and its Effects, Calonge, Spain, 1991.

Gipps, C., McCallum, B., McAlister, S. & Brown, M., 1991, 'National assessment at seven; some emerging themes', paper presented to the Annual Conference of the British Educational Research Association, Nottingham, 1991.

Graded Assessment in Mathematics (GAIM). 1988, *GAIM Development Pack*, Macmillan, London.

Inner London Education Authority: 1976, *Checkpoints*, Inner London Education Authority Learning Resources Scheme, London. (This is soon to be republished in a revised version by Harcourt Brace Janovich).

Love, E. and Shiu, C.: 1991, 'Criteria and meaning in teacher assessment of mathematics', in Pre-proceedings of ICME Conference Assessment in Mathematics Education and its Effects, Calonge, Spain, 1991.

Robitaille, D.F. (Ed.), 1989, *Evaluation and Assessment in Mathematics*, Science and Technology Education Series No. 32, UNESCO, Paris.

School Examinations and Assessment Council (SEAC): 1991, *Teacher Assessment at Key Stage 3: Mathematics*, SEAC, London.

School Mathematics Project (SMP): 1989, *Open Ended Tasks*, School Mathematics Project, Southampton.

Secondary Examinations Council: 1985, Report of a Working party - Mathematics - Draft Grade Criteria, Secondary Examinations Council, London.

Shell Centre for Mathematical Education/Joint Matriculation Board: 1984-90, Teachers packs in the series *Teaching Strategic Skills* and *Numeracy through Problem Solving*, Longman, London.

Travers, K.J.: 1989, Curriculum-linked assessment: a model based on the Second International Mathematics Study, in Robitaille, D.F. (Ed.), *Evaluation and Assessment in Mathematics*, Science and Technology Education Series No. 32, UNESCO, Paris

Margaret Brown
King's College London,
United Kingdom

CHRIS LITTLE

THE SCHOOL MATHEMATICS PROJECT: SOME SECONDARY SCHOOL ASSESSMENT INITIATIVES IN ENGLAND

1. INTRODUCTION

The established English tradition (see, for example Howson, 1982) of public examinations at ages 16+ and 18+ has had an enormous effect on the implementation of the mathematics curriculum in English secondary schools.

To be effective, all secondary school curriculum development must be underscored by suitable examinations, and the *School Mathematics Project (SMP)* has from its early days worked with examining bodies to ensure that public examinations are developed which reflect and support curricular aims.

This paper outlines recent developmental work on assessment conducted by SMP in England and Wales, and discusses some of its implications. The references SMP 1, SMP 2, etc. refer to examination syllabuses listed in the References.

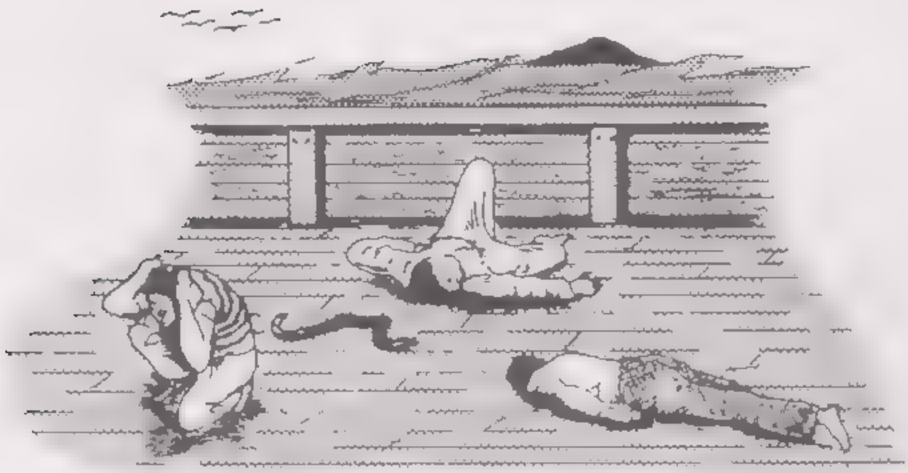
2. DEVELOPMENTS IN WRITTEN EXAMINATIONS

Context

Although the range of mathematical problem solving which can be tested in a timed written examination is restricted, there is scope even in this medium to set questions which reflect curriculum goals, such as embedding mathematics in contexts which are concrete and meaningful to children. For example, the question in Figure 1 (from SMP 1) is psychologically quite different to "calculate 54% of 1955", even though their mathematical solution is identical. Of course, not all mathematical topics (for example, prime numbers) can or indeed should be treated in this way.

Finding and using contexts which are apposite and readily assimilated by candidates is not an easy task. SMP uses teams of teachers to pool ideas for questions, rather than relying on the ingenuity of a single examiner.

M. Niss (ed.), Cases of Assessment in Mathematics Education, 85–97.



In the eighteenth century Captain Anson sailed round the world He started out with a crew of 1955 During the voyage 54% of the crew died of fever.
How many died during the trip?

Figure 1

Differentiated Papers

The *Cockcroft Report* (DES, 1982) highlighted the demoralizing effect of examinations which grade candidates on the basis of failure This has led to the development of *differentiated papers*, in which a single syllabus and examination is replaced by a set of syllabuses and papers, of increasing depth and difficulty, from which candidates select according to their aptitude. Depending on the papers taken, a restricted range of grades is then available.

SMP's current *GCSE* (*General Certificate of Secondary Education*) syllabus (SMP 1) is an example of a scheme of this type, in which pupils select two from a ladder of four written papers, each of which has its own syllabus. A restricted range of pass grades, which run from A to G, are available for each level of entry (see Table 1).

For each level, the mark range for the award of grades is approximately 40% to 80%, and grading is therefore soundly based on positive achievement rather than failure. SMP has also extended the use of differentiated papers to one of its 18+ *GCE* (*General Certificate of Education*) A Level syllabuses (SMP 2).

As with contextualisation, differentiated schemes of assessment have added to the complexity of the setting, marking and awarding of grades. Another difficulty is that some candidates are ungraded at the Higher Level through poor performance, who would easily merit a grade had they been entered at a lower level. This places the onus on teachers to enter candidates at the correct level, and students themselves (and their parents!)

Level	Papers	Grades available
Foundation	1 and 2	E, F, G
Intermediate	2 and 3	C, D, E, F
Higher	3 and 4	A, B, C, D

Table 1 *SMP GCSE scheme*

must accept that they have a restricted range of grades available.

Nevertheless, differentiated papers have undoubtedly succeeded in making written end-of-course examinations a more positive experience for students.

Comprehension Papers

The curricular "backwash" from timed written examinations need not of necessity be educationally damaging. An alternative type of written paper which has been developed by SMP in its Advanced Level syllabuses (SMP 3 and SMP 4) is the *comprehensive paper*. In this, candidates study mathematical articles, and then answer questions which either test comprehension of the mathematics, or ask them to expand or develop the ideas further.

Unlike conventional written papers, which tends to close down and focus the student on the syllabus content, the effect of comprehension papers has been to open out the mathematics studied, improve the students' ability to read mathematics intelligently, and broaden their attitudes to mathematics. This type of paper has also proved to be an effective discriminator of performance.

3. COURSEWORK ASSESSMENT

In England, *coursework assessment* in the GCSE has been used as a vehicle for broadening classroom practice to include discussion, investigative and practical work, as advocated in the influential Cockcroft Report (DES, 1982).

For mathematics teachers used to the comfortable world of "right" and "wrong" answers, the problems of assessing extended investigative work, or oral discussion of mathematics, are considerable. More qualitative assessment of tasks also poses problems of the standardization of marking or grading needed in a public examination.

Coursework Tasks

Coursework development poses a number of questions:

- What *types of task* are to be used, and who selects the task (the pupil, the teacher, the school department, or the external examining body)?
- What *conditions* are laid down for conducting the work, and how long is allowed to complete the tasks?
- How are the tasks to be *assessed*?
- What *weight* is given to the tasks in the overall scheme of assessment?

In its development work for SMP 1, SMP has moved from relatively well defined prescribed tasks, varying in length from one hour to two weeks' work in mathematics, with task-specific mark schemes, towards defining broader categories of extended open-ended tasks, selected by the teacher, and marked in relation to general process criteria.

The original prescribed tasks fell into five categories of work: drawing, geometrical pattern, investigations, statistical survey, and sampling. Some tasks were extended practical tasks, such as the design of a dog kennel, others were shorter tasks, designed to be done in about one hour. The tasks were accompanied by mark schemes, which were developed on the basis of trial pupil scripts (see Figure 2).

Detailed administration guidance was given to schools, to ensure "fairness". Thus, pupils were not allowed to work cooperatively, and tasks were to be given "cold" without any introduction or advice from the teacher. Teachers were not allowed to report marks to candidates, or to give specific feedback on the tasks. Teachers were encouraged to meet and discuss the interpretation of mark schemes, in order to standardize their marking. Although coursework assessment was new to nearly all the teachers, most were able to cope with the new demands placed on them.

However, prescribing the tasks, prohibiting feedback, and effectively prescribing the responses to the tasks by issuing mark schemes, all limit the educational worth of this type of coursework. SMP therefore sought to develop coursework tasks, called Open-Ended Tasks (OETs) which were more genuinely open-ended and less restrictive.

SMP 1 now defines two broad categories of OETs:

(a) *Practical or applied work*, in which pupils apply mathematics to real life problems;
(b) *Investigational tasks*, in which pupils explore intrinsically *mathematical* problems.

1990 RESIT ENTRY ONLY.

London East Anglian Grou
Midland Examining Grou
On behalf of Groups nationall

SMP 11 — 16 Coursework Task

CIRCLES, DOTS AND LINES

Level: Foundation

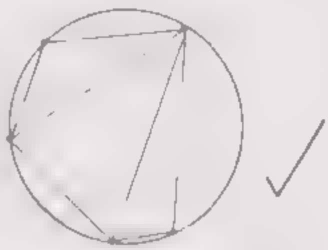
Category: ST *(In school under supervision, time limit 1 hour)*

REMEMBER. Show all your working clearly so that someone else can follow what you did.

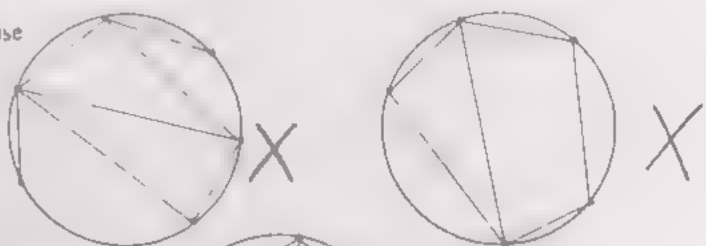
In this task you are going to investigate lines joining dots on a circle

You must join as many dots as you can, but lines must not cross

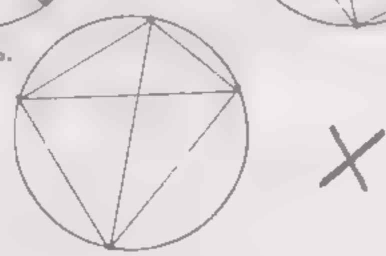
In this diagram as many lines as possible have been drawn.



These diagrams are not allowed because you could draw more lines.



This diagram is not allowed because some lines cross.



1. *(a)* Use some of the circles on the worksheet to draw more diagrams.

   (b) Investigate the connection between the number of dots and the number of triangles.

   (c) How many triangles would there be if you had 79 dots?
   Describe clearly how you got your answer.

2. Now look at your diagrams again. This time you are considering the number of lines.

   (a) Investigate the connection between the number of dots and the number of lines.

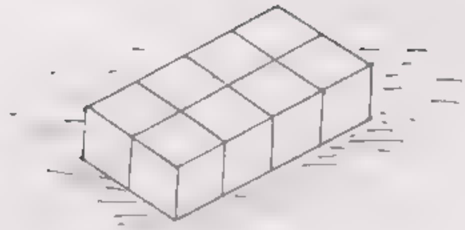
   (b) How many lines would there be if you had 113 dots?
   Describe clearly how you got your answer.

Figure 2

Examples of OETs

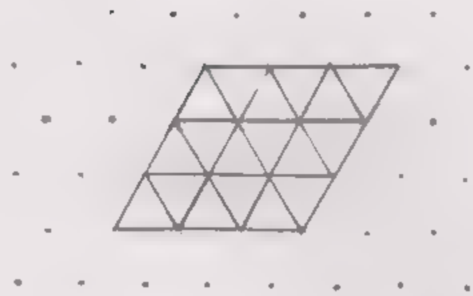
Investigational

1. Hidden Faces



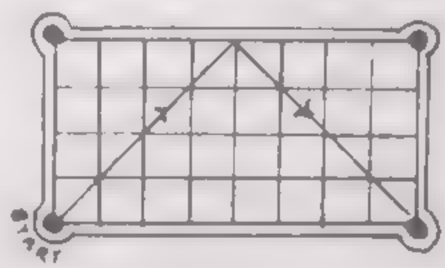
When cubes are placed together on a surface, it is impossible to see some of the faces of the cubes. Investigate the hidden faces.

2. Growing Rhombuses



How many rhombuses are there on this shape? Investigate this type of diagram.

3. The Strange Billiard Table



This billiard table is a little odd. It only has four pockets, and the base is divided into squares. Only one billiard ball is used, and it is always struck from the corner at 45 degrees to the side. (The ball also rebounds at 45 degrees to the side). Investigate what happens for tables of different sizes.

Practical

1. Car Park

It has been suggested that when the temporary huts in a certain area are removed, the area is turned into a car park to replace one of the existing ones. Investigate this.

2. Smarties

You can buy Smarties in different types of package. Design a new package.

3. Shopping

Where does your family shop for food? Is this a sensible decision?

Figure 3 *Examples of OETs*

Although SMP now provides some examples of each sort (see Figure 3), the onus of finding suitable starting points for OETs is placed on schools. This enables tasks to reflect local interests, and reduces the possibility of project work and "solutions" becoming standardized. Greater freedom can be allowed for pupils to collaborate with their teachers and fellow pupils, provided the resulting work can be assessed fairly.

Rather than using mark schemes specific to the task, the work done is assessed with reference to general criteria which characterize the investigative processes involved.

For *Practical OETs*, these categories are:

Identifying:	Analyzing, Planning
Implementing:	Modeling, Experimenting/Questioning, Sampling/Collecting, Measuring, Processing Data, Representing, Checking/-Optimizing
Reviewing:	Formulating a Solution, Communicating, Interpreting, Adapting

For *Investigational OETs*:

Identifying:	Questioning/Extending, Planning, Getting Started/Simplifying
Implementing:	Working Systematically, Classifying, Symbolizing/Recording, Conjecturing/Generalizing, Checking/Proving
Reviewing:	Summarizing, Communicating, Extending

The assessment sheets developed use a "thermometer" approach rather than assigning marks or grades for each category. The assessments on each "thermometer" are then aggregated, by eye or judgement rather than by an arithmetically defined procedure, to give an overall grade for the piece of work. In-service training materials have been produced, after substantial trialing, which give detailed descriptions of the processes, together with annotated student scripts.

Assessing work in this way is quite different to using coursework tasks with task-specific mark schemes. Here, the assessment scheme is less objective, and providing a framework within which teachers can apply their own professional judgements to establish the worth of the work. Experience suggests that while the assessment of the teachers on the individual criteria varies substantially, the overall grade awarded shows good agreement. It is tempting to deduce from this that an overall holistic assessment of the worth of a piece of investigative work would be quicker and no less reliable! But this procedure gives little help when it comes to standardizing gradings, since it is probable that teachers will each use different criteria, such as quantity of work, effort, accuracy of calculations, and so on. Applying detailed criteria, although daunting at first, becomes easier with practice, after which they become more familiar.

The importance of GCE Advanced Level examinations for university matriculation has perhaps hindered the acceptance of these rich but intrinsically less consistent assessment methods at this level. Nevertheless, SMP has introduced project assessment in its Advanced Level 18+

syllabuses (SMP 3 and 4). The work has focussed more on mathematical content than at GCSE level. SMP 4 contains a compulsory *Problem Solving Module*, which is assessed through a one-hour comprehension test, and two pieces of problem-solving work. The criteria for assessment are described through the categories Design, Mathematical Enquiry, Rigor, Model Formulation, Interpretation/Validation, Initiative, Content and Communication.

Assessment of Mental Skills

Calculators are freely used in UK written examinations. Most GCSE Mathematics syllabuses therefore use mental tests to assess other methods of computation. These tests have proved to be quick, straightforward and painless to administer.

The tests need not to be restricted to the "traditional" number work. SMP has also used them in SMP 5 to test estimation, spatial visualization, and even algebra. For example:

- Sketch a hexagon which has exactly two lines of symmetry.
- Estimate the value of sine one hundred degrees.
- The point two comma negative three is reflected in the x-axis. What are the coordinates of the image?
- A formula for the perimeter of a semi-circle is $\pi r + 2r$. Factorize this expression.

An issue which has prompted much discussion is whether pupils should be allowed to use working, or make notes of questions. The "purist" approach here is to insist on candidates writing nothing but the answer, in order to "force" them to process the information mentally. The "pragmatist" approach is to allow some working on the grounds that it is artificial to deprive pupils of these methods.

Oral Assessment

As discussion is a vital component of mathematical activity, there has been considerable interest in assessing this. There are two aspects:

- How can we assess the mathematics pupils know using oral methods?
- How can we assess the pupils' ability to express mathematics orally?

SMP has used two approaches. In SMP 1, "communication" is assessed as one of the process criteria for open project work. Alternatively, in SMP 5 a more formal scripted interview has been used (see Figure 4). This has the benefit that students (and teachers) take oral assessment more seriously.

Interviews also create breathing space within the classroom for teachers to focus their attention on oral aspects.

Interviews have been enjoyed by most students, and have furnished teachers with rich insights into the depth of their understanding. However, as teachers need to spend at least fifteen minutes with each student, it is extremely time-consuming, and this has made it unrealistic for most schools. Oral assessment is moreover exceptionally difficult to standardize. Even more doubtful is the feasibility of making robust assessments of collaborative discussion between pupils.

Graded Assessment

The aspects of coursework dealt with thus far are designed to assess skills and processes for which written examination papers are not appropriate. However, much of the ongoing class activity of pupils consists of work of this kind, that is answering written questions or exercises. By excluding the assessment of these more routine aspects of class work from coursework assessment, one is neglecting a large source of information. Moreover, if regular classroom assessments are credited to pupils as part of their GCSE assessments, then this can act as a strong motivator, provided the assessments are at an appropriate level of difficulty.

In the SMP Graduated Assessment Scheme (SMP 6), pupils sit regular written tests, or 'Recaps', based on the curriculum material they are studying (see Figure 5). To pass a Recap, they need to achieve at least 80% of the marks for questions in each of the categories Using Arithmetic, Interpreting Data, Applying Spatial Skills, and Interpreting Three Dimensions. If they fail to reach this target, then after further study they may resit the Recap, using a parallel version. In addition to these written tests, at each stage, pupils take one-to-one, scripted, oral and practical tests, and mental and estimation tests. Pupils may also submit extended, more sustained topic work. On completion of each stage, pupils receive a *Stage Certificate* which lists their achievements in some detail.

Most of the marking involved in the scheme is straightforward. However, class sizes have to be small in order to cope with the large amount of one-to-one assessment and administration involved.

Many of the low-attaining students who use this scheme in the past have been classified as 'failures', and frequently truanted from mathematics classes before leaving school at 16. The short-term goals provided by the staged assessments have proved to be a powerful motivator.

The test-remediate-retest cycle implied by graded assessment has also affected teaching methods, which have become more diagnostic, selecting specific mathematical tasks to help individual difficulties.

Oral Interview Script (Foundation Level)
1991 Entry

You will need

Prepared item diagrams (tile patterns)
Unprepared item cards (door designs)
Reference sheet for unprepared item (door designs)
Red and white tiles
13th card (door with triangular pattern)

Script

A prepared item

Show candidate sheet with diagrams A and B. Point out that the shaded squares in the diagrams stand for red tiles. Then choose diagram A or diagram B.

Suppose I make up the sixth pattern of diagram A/B with these tiles. How many red and white tiles would I need?

Now, without showing me the diagram tell me how you make up the sixth pattern.

Follow instructions precisely, making deliberate "mistakes" to prompt for greater precision. Keep oral prompts short, e.g. "like this ...?"

Marking

No marks to be awarded here. If answer is incorrect prompt with "are you sure?" If still, wrong, ask about fifth pattern, then repeat question for sixth. [Answers: A needs 6 red (shaded) and 10 white, B needs 6 red (shaded) and 18 white].

Answer

3 Fluent and accurate instructions, no verbal prompting needed
2 Instructions succesfull, but lacking a little precision, or some verbal prompting needed
1 Succeeded with instructions, but vague, and substantial prompting needed
0 Fails to describe pattern, even after prompting

DIAGRAM A

DIAGRAM B

Figure 4 *Oral interview script*

SMP Graduated Assessment
REVISED

Stage 3 Recap A

Student's Marks	UA 20	IO 8	AS 3		Marked by	

Name ______ Group ______ Date ______

UA 1 Here are two bars of chocolate

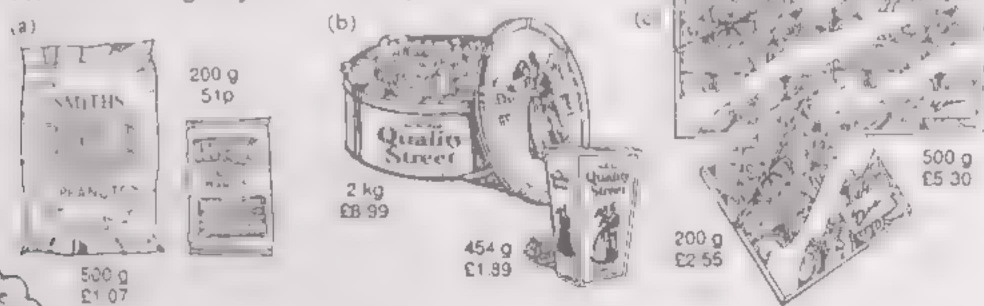
(a) About how many grams do you get for 1p in the large size? ______ g

(b) About how many grams do you get for 1p in the small size? ______ g

(c) Which size gives you more for your money, the large or the small? ______

Write large or small

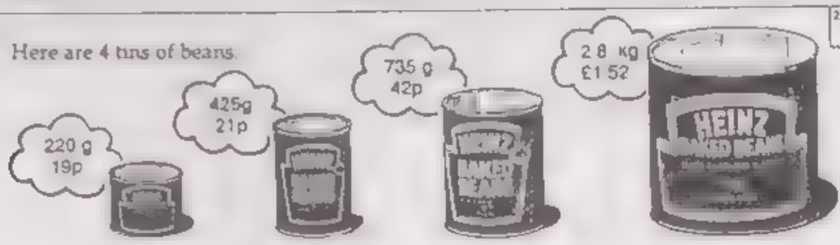
UA 2 Here are some tins and packets with their prices
Decide which size gives you the most for your money.



Write large or small

(a) ______ (b) ______ (c) ______

UA 3 Here are 4 tins of beans.



Which tin gives you the most for your money? ______

 1 Page total

Figure 5

4. IMPLEMENTATION

Implementing new assessment methods successfully is costly. Commitments need to be made to the training of teachers and moderators to implement and monitor the assessment procedures.

In the case of the current SMP 1, which had a candidature of 160,000 in 1991, well over 50 in-service courses have been run for teachers on the new assessment methods, and over 100 coursework moderators were appointed to monitor the examination this year. In order to be accredited to administer the Graduated Assessment Scheme, which has a current candidature of 65,000, teachers have to attend a compulsory two day in-service course, and to date over 5,000 teachers have attended such courses.

All the courses run have been based on the practical experience of teachers who have helped to develop or pilot the assessment methods.

Successful implementation depends on piloting new ideas on a small scale, introducing new ideas gradually, and providing adequate support to teachers. It is important to learn to walk before you run. Although the prescribed coursework tasks described in Section 3 have educational limitations, they provide a valuable learning experience for teachers who are not used to assessing open project work. OETs have replaced these gradually, initially on an optional basis, so that schools with different levels of expertise and experience can select coursework which suits their needs.

5. CONCLUSION

SMP's experience suggests that it is possible to broaden the range of assessment methods used in large-scale public examinations, and that this can have a beneficial effect on the school mathematics curriculum.

The various assessment methods discussed suggest the existence of an Inverse Law of Assessment: the reliability of the evidence obtained is inversely proportional to its educational validity! If public examinations are to reflect and promote a dynamic, creative and intellectually stimulating mathematics curriculum, then they must be prepared to use assessment tools which rely more on the professional judgement of teachers, and less on objective externally devised tests.

Before extending the range and variety of assessment methods, it is therefore crucial to consider how they will be successfully implemented by teachers. It is salutory to note that at the time of writing, the weighting given to coursework assessment in the GCSE is to be reduced, because of the politically perceived unreliability of the GCSE. It is important to weigh the educational benefits of subjective, teacher-dependent assessment against society's requirements to have "fair", consistent, and objective public examinations.

REFERENCES

(a) SMP Examination syllabuses

SMP 1 General Certificate of Secondary Education, Mathematics (*SMP 11–16*), syllabus code 1653, National Curriculum Project, administered by Midland Examining Group and University of London Examinations and Assessment Council

SMP 2 General Certificate of Education, Advanced level, *SMP Mathematics*, administered by Oxford and Cambridge Schools Examinations Board

SMP 3 General Certificate of Education, Advanced level, *SMP Further Mathematics*, administered by Oxford and Cambridge Schools Examinations Board

SMP 4 General Certificate of Education, Advanced level, *16–19 Mathematics*, administered by the Joint Matriculation Board

SMP 5 General Certificate of Secondary Education, *Mathematics (SMP)*, syllabus code 7451, Mode 2 syllabus administered by Midland Examining Group (1988–1991)

SMP 6 *SMP Graduated Assessment Scheme*, administered by Oxford and Cambridge Schools Examinations Board

(b) Other references

Department of Education and Science (DES)/Welsh Office, Committee of Inquiring into the Teaching of Mathematics in Schools 1982, *Mathematics Counts*, ("The Cockcroft Report"), Her Majesty's Stationery Office, London.

Howson, A G 1982, *A History of Mathematics Education in England*, Cambridge University Press.

Chris Little
The School Mathematics Project,
University of Southampton,
United Kingdom

LUCIANA BAZZINI

THE TEACHING/LEARNING PROCESS AND ASSESSMENT PRACTICE: TWO INTERTWINED SIDES OF MATHEMATICS EDUCATION

1. INTRODUCTION

This contribution to the ICMI Study on Assessment in Mathematics Education and its Effects deals with the relationship between instruction and assessment and views their reciprocal dependence. A strong emphasis on the interconnection between teaching, learning, and assessing is found in the so-called *dynamic assessment* that derives from the analysis of the Vygotskian *zone of proximal development* and, therefore, is highly concerned with the individuals' responsiveness to teaching. An approach to assessment which considers evaluation procedures as continually intertwined with teaching/learning procedures is widely held in the Italian system. It is also deeply rooted in the Italian tradition. In the second part of the paper some main features of the Italian assessment system are outlined. Finally, an example of how the assessment problem is faced in a study on curriculum development in primary school is given.

2. ASSESSMENT AND INSTRUCTION

In the last decades, some changing views on mathematics education have led to an increasing concern for the role of the individual's own activity within the teaching/learning process. As Christiansen and Walther (1986) point out, three tendencies may be noted. A growing acceptance of the view that a prerequisite for "meaningful" learning of any part of school mathematics is the individual's personal involvement and reflection, an emphasis not only on the results of the mathematical working process, but also on the working process itself, and finally a tendency to see the teaching of school mathematics not only as instruction, but as a long-term process of interaction. According to these trends, and strongly contrasting with a traditional view of learning in which the child moves through a sequence of increasingly difficult tasks, the assessment methodology called dynamic assessment provides new perspectives.

M. Niss (ed.), Cases of Assessment in Mathematics Education, 99–106.

Dynamic assessment refers to the assessment of individuals' responsiveness to teaching (Feuerstein, Rand & Hoffmann, 1979) or zone of sensitivity to instruction (Vygotsky, 1978). As Brandford et al. (1989) notice, the methods of assessment are different from those used in standardized, static assessments, such as intelligence tests and achievement tests. One main feature of dynamic assessment is a systematic attempt to actively change various components of tasks and approaches to teaching in order to find the conditions that are most effective for each child. There is a strong relation with Vygotsky's notion of zone of proximal development. In his view, every specific state of the child's development is characterized by the *actual* developmental level and the level of *potential* development.

"The zone of proximal development is the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance, or in collaboration with more capable peers" (Vygotsky, 1978, p. 86).

The learner progresses in the zone of proximal development by means of educational guidance and support. In this perspective, there is no dichotomy between the learner's independence and the guidance provided by the teacher: The two aspects, the learner's autonomy and the educational support, are interdependent. Various forms of knowledge cannot be developed spontaneously by the learner, but must be mediated by educational forms of support. One main interest in the zone of proximal development is the opportunity to observe the child's assisted progress and do on-line diagnosis. In contrast, when a child's competence is assessed on some static, independent test, only the child's actual level of development is reflected.

As Newman, Griffin & Cole (1989) point out, the two conceptions (the traditional, static assessment and the dynamic assessment)

lead to very different approaches to monitoring the child's progress and assessing his or her abilities. In the traditional view, competence is measured by successful performance of a task at a particular point in the sequence. Change over time is seen in improved performance of a task or in movement up the sequence. In either case, the child's individual performance is assessed. The zone of proximal development provides a strikingly different approach. Instead of giving the children a task and measuring how well they do or how badly they do, one can give the children the task and observe how much and what kind of help they need in order to complete the task successfully. In this approach, the child is not assessed alone. Rather, the social system of the teacher and the child is dynamically assessed to determine how far along it had progressed" (Newman et al., p. 77).

3. THE ASSESSMENT SYSTEM IN ITALY

Coming to consider the peculiarities of the assessment system in Italy, we can say that the main trend is oriented towards methods which take into

account the child's cognitive and affective development in its entirety. This is as true today as it was in the past. More precisely, a close connection between teaching and testing is evident, in the sense that they proceed together in the educational process, the latter as a means for the former. Although the teaching system is quite varied with regard to types of school and pupils' age-levels, we can recognize the tendency to be in accordance with the dynamic assessment methodology, at least in compulsory education (ages 6–14). I will return to this point later.

We must also recognize that most Italian researchers are more interested in evaluating programs and curricula than assessing individual students. However, since the focus of this study is on pupil assessment and not on curriculum evaluation, I will limit myself to the former aspect. In Italy, the teaching system is centralized, in the sense that the Ministry of Education takes the responsibility of issuing programs for every level of pre-university schools. Consequently, the government programs play a crucial role in orienting didactic practice. Generally speaking, we can say that the programs are guidelines concerning educational objectives, contents, and methods. They are not issued frequently. The last issue of programs for primary school (age 6–11) was in 1985, for lower secondary school in 1979, and for upper secondary school in 1990 (these latter are just experimental programs). We have the same programs for the age levels up to 14 years; for the different kinds of upper secondary schools there are specific programs.

As far as mathematics is concerned, the programs present an image of coherence: At every level of the school, they recognize that the main objective of mathematics education is to train pupils in the approaching of and solving of problems, in making suitable representations, and in interpreting and verifying results. Doing mathematics is seen as a systematic and progressive activity which starts in the earliest grades and proceeds in a spiral or fan-shaped development. The student is encouraged to be active in the construction of his/her knowledge. But this does not mean that the practical teaching is always in accordance with the programs' guidelines.

In the programs, the assessment problem is contemplated explicitly. For example, the government programs for primary schools point out that in order to secure an effective evaluation of the starting point and the arrival point of the child's learning processes and difficulties, and in order to get a good individual and collective comparison, primary school teachers must systematically collect information on a child's cognitive and affective development. Different ways of collecting data should be used. Objective tests and other informal kinds of assessment are suitable. In short, we can see an orientation towards a variety of assessment methods, in accordance with and as a consequence of a multifaceted approach to mathematical activities.

From a strictly formal point of view, the teaching system provides three *assessment events*: At the end of primary school, at the end of lower

secondary school, and at the end of upper secondary school. A commission formed of the teachers of the class and two other teachers of the school assesses pupils at the end of primary school, on the basis of two written examinations and an interview. At the end of lower secondary school, the students are assessed by a commission consisting of the teachers of the school and an outside member, appointed by the central authority, who acts as chair. The commission assesses the students by means of written tasks (on language, mathematics, and foreign language) and a discussion.

During the whole period of compulsory education, no score is used. Pupils are assessed by means of an individual judgement concerning profit and behavior. At the end of upper secondary school a commission of teachers, coming from outside and appointed by the Ministry of Education, assesses the student's achievements on the basis of two written tasks, whose specific content is established by the Ministry, and by a discussion based on four subjects, two chosen by the commission and two by the student. The final score is a composite of the student's profit in all the disciplines. This final examination is a powerful landmark in orienting classroom practice in the last years of secondary school. In contrast, since the examinations at the end of primary and lower secondary school are established by the teachers of the school, they do not constitute a forced constraint on previous activities. It is evident that there is a greater autonomy for the teachers in compulsory education than for teachers at a more advanced level of education.

4. LARGE-SCALE INVESTIGATION

The issue of new programs usually implies a phase of transition and innovational ferment. On one hand, the government programs feel the influence of the work that various didactic research groups (the so-called *Nuclei de Ricerca Didattica*) have been carrying out for several years; on the other hand, the work acts as a stimulus towards innovation. In the phase of transition, the need of knowing the reality is of great interest. This need to analyze the real situation in which the schools find themselves leads to some large-scale testing.

The large-scale tests are not aimed at assessing students, but rather at assessing the real state of things in schools. Standardized tests are quite in contrast with our traditional approach to assessment. In fact, teachers have always been reluctant to adopt them and, in particular, the multiple-choice ones. This is mainly due to the deep-rooted idea that knowledge of the subject by the student cannot be well assessed by standardized tests. This notwithstanding, the large scale-tests provide useful information. Let me recall two examples. The first concerns the so-called *VAMIO* study (Verifica delle Abilità Matematiche nella Scuola dell'Obbligo), which was supported by the *CEDE* (Centro Europeo dell'Educazione) in 1985/86. It

was based on the methods of the IEA surveys and was aimed at producing a standardized test to assess pupils at the end of lower secondary school (Bolletta, 1987). The second example concerns a study organized in 1985 by the *IRRSAE Lombardia* (Istituto Regionale di Ricerca, Sperimentazione e Aggiornamento Educativi). This study was aimed at verifying the state of things in the passage from primary school to lower secondary school (Bazzini, 1989a) and from lower secondary school to upper secondary school (Reggiani, 1989). I was responsible for the development of the test of achievement in mathematics at the end of primary school in this study. The test consisted of multiple-choice items and, therefore, did not suit the usual models of assessment. The test was given to 1,500 students at the beginning of the first grade of lower secondary school of the 1986/87 school year. The students were randomly chosen from the population of students attending public schools in Lombardia. Without going into detail, I would like to point out just some particular facts. There emerged from the results a picture of students as more concerned with arithmetic computation than with problem solving, although arithmetic is traditionally embedded in problems. Good performances in computation do not often correspond to the capacity of mathematizing a given situation or choosing the right operation to solve a problem. We also noticed that the ability to continue a given sequence of numbers or to discover a regularity, which was not explicitly evident, did not seem to be in the baggage of the average student. This information was vital to us in successive research.

5. OTHER KINDS OF ASSESSMENT

With the growing interest in the teaching/learning processes, the necessity of an accurate analysis of the assessment problem was a consequence. The question has been considered by Bartolini Bussi (1989) in the framework of a study on social interaction in the classroom. A more technical analysis of the assessment instruments is given by Guala (1989). The debate is still open.

Generally speaking, we can say that, in primary education, the most common approach is in accordance with dynamic assessment, as we have already observed. There is constant attention to observing what children are able to do and how they act. On the grounds of this, the teacher is ready to adjust his/her intervention in order to fit it with the student's performance. We can also recognize an agreement with what has been called *informal assessment* (Clarke, Clarke & Lovitt, 1990). Informal assessment means a collection of assessment information coinciding with instruction, that is sensitive to process as well as product. By contrast, *formal assessment* requires the organization of an *assessment event*.

In secondary education, the use of formal assessment increases, although the informal tools (classroom discussion, interviews, free observation, etc.)

are not abandoned. Moreover, assessment must continue in time, since instruction is a long-term process. When the same teacher teaches the same class for more than one year, as is usual in our system, a continuous observation of the student's development is possible and suitable. This is a very relevant feature of assessment, because it keeps an account of cognitive as well as affective domains.

The relevance of continuous assessment in primary education has been recently stressed by Webb & Briars (1990). These authors start from the basic consideration that mathematics is a dynamic, interconnected system and students' knowledge of mathematical concepts and procedures, problem solving, and reasoning develop and mature over a period of years. The knowledge of the meanings students assign to the mathematical ideas they are learning is very important and assessment, then, must be an interaction between teacher and student, with the teacher continually seeking to understand what a student can do and how a student is able to do it, and then using this information to guide instruction (Webb & Briars, 1990, p. 108).

6. THE APPROACH TO ASSESSMENT IN A STUDY ON CURRICULUM DEVELOPMENT IN PRIMARY SCHOOL

I now focus on the basic assumptions adopted in a study on the development of the mathematical curriculum in primary school. This study, which the Nucleo de Ricerca Didattica of Pavia, carried out for several years, is aimed at putting into practice the spirit and suggestions of the government programs, which are widely shared as far as methods and contents are concerned. A close cooperation between university researchers and school teachers is a particular feature of the study. As far as assessment is concerned, the study is embedded in the cultural atmosphere I have tried to describe.

From a general standpoint, we can identify two main streams: One concerns a quantitative analysis of test results, while the other is concerned with a qualitative analysis of pupils' behavior. The former is more oriented to curriculum evaluation, the latter to pupils' assessment (Bazzini, 1989b). Following the programs' recommendations, data on children's progress are systematically collected. The instruments used are of a different nature: Written tests, classroom discussions, individual interviews, and free observation. Some written tests are established by the entire group involved in the study, and they are equally administered to all pupils at the end of the first term and at the end of the school year. They are not multiple-choice tests. The tests are conceived in accordance with the tasks pupils are used to do in the classroom. Nevertheless, in some cases, pupils recognize them as a means of control. For each item, the percentage of correct answers is calculated; this gives information about the effects of the

work which was carried out. The results of each test provide us with the state of things at a certain moment. When compared with the results of previous and successive tests, they can also give an idea of the way things are going. This kind of information is very useful in the planning of successive activities.

Together with the written test, several informal tools are used to assess pupils. These informal methods are not only a means of assessment, but an integral part of instruction. As already observed, teaching, learning, and assessing present a continuous interdependence, whose importance is also stressed by Marshall (1989) in her effort to shift from assessment procedures in problem solving based upon statistical and psychometric models to procedures based upon cognitive models of learning and memory.

Last, but not least, is the problem concerning the teacher's capacity to observe pupils' reactions. In many ways, the teacher is more like a cognitive researcher than a tester: Surely, the teacher's competence in cognitive processes is a fundamental basis for understanding pupils' performances and assessing them. To this purpose, we usually devote time and energy discussing and analyzing students' protocols, and the different strategies used in solving problems. Particular emphasis is given to recurring errors and to finding the source of these errors. The study in question has led to a growing awareness, on the part of the teacher, of his/her fundamental role in identifying pupils' knowledge, and of the need to clarify some pupils' behaviors which previously may have seemed incomprehensible.

Teachers' capacity to notice and interprete classroom movements plays a crucial role in the teaching and learning process. Silver & Kilpatrick (1989) observed that many aspects of problem-solving performance seem likely to elude efforts to improve testing through technique and technology. Their assessment requires the skills of a sensitive, informed teacher. The teacher who can conduct a problem-solving lesson can also assess how student have responded to it and how their performance has improved as a consequence. What is needed are reskilled teachers who are able to construct their own assessment instruments and to determine what and how their students are doing when they face mathematical questions.

REFERENCES

Bartolini Bussi, M. 1989, 'Evaluation of teaching sequences that include individual and collective activities. Two case studies', in *Proceedings of the First Italian German Bilateral Symposium on Didactics of Mathematics*, TID, CNR, Progetto.

Bazzini, L. 1989a, 'From primary to secondary education. An investigation of pupils mathematical abilities' *Proceedings of the 41 CIEAEM Conference*, Bruxelles.

Bazzini, L.: 1989b, 'Curriculum development and evaluation in primary school' *Proceedings of the First Italian-German Bilateral Symposium on Didactics of Mathematics*, TID, CNR, Progetto.

Bolletta, R.: 1987, 'The real curriculum in "Scuola Media" after the new programs of 1979 in Italy', paper presented to the International Conference on Comparative Studies of Mathematical Curricula in Different Countries, CEDE, Frascati.

Bransford, J., Hasselbring, T., Barron, B., Kulewicz, S., Littlefield, J. & Goin, L.: 1989, 'Uses of macro-contexts to facilitate mathematical thinking', in *The Teaching and Assessing of Mathematical Problem Solving*, Erlbaum, NCTM, Hillsdale, NJ.

Christiansen, B. & Walther, G.: 1986, 'Task and activity', in Christiansen, B., Howson, A.G. & Otte, M. (Eds.), *Perspectives on mathematics education*, Reidel Publishing Co., Dordrecht.

Clarke, D.J., Clarke, D.M. & Lovitt, C.J.: 1990, 'Changes in mathematics teaching call for assessment alternatives', in *Teaching and Learning Mathematics in the 1990s. NCTM Yearbook*, NCTM, Reston, VA.

Feuerstein, R., Rand, Y. & Hoffman, M.: 1979, *The dynamic assessment of retarded performers. The learning potential assessment device: Theory, instruments, and techniques*, University Park Press, Baltimore.

Guala, E.: 1989, 'Mathematical learning in secondary school. Some evaluation methodologies' *Proceedings of the First Italian-German Bilateral Symposium on Didactics of Mathematics*, TID, CNR, Progetto.

Marshall, S.P.: 1989, 'Assessing problem solving. A short term remedy and a long term solution', in *The Teaching and Assessing of Mathematical Problem Solving*, Erlbaum, NCTM, Hillsdale, NJ.

Newman, D., Griffin, P. & Cole, M.: 1989, *The construction zone. Working for cognitive change in school*, Cambridge University Press, Cambridge, MS.

Reggiani, M.: 1989, 'Test of the learning of mathematics at the end of the obligatory school (14 years old)', in *Proceedings of the 41 CIEAEM Conference*, Bruxelles.

Silver, E.A. & Kilpatrick, J.: 1989, 'Testing mathematical problem solving', *The teaching and assessing of mathematical problem solving*, Erlbaum, NCTM, Hillside, NJ.

Vygotsky, L.S.: 1978, *Mind in society. The development of higher psychological processes*, Harvard University Press, Cambridge, MA.

Webb, N. & Briars, D.: 1990, 'Assessment in mathematics classroom, K-8', *Teaching and Learning Mathematics in the 1990s. NCTM Yearbook*, NCTM, Reston, VA.

Luciana Bazzini
Department of Mathematics
University of Pavia
Italy

GUNNAR GJONE

TYPES OF PROBLEMS AND HOW STUDENTS IN NORWAY SOLVE THEM

1. INTRODUCTION: THE NORWEGIAN SYSTEM OF ASSESSMENT IN MATHEMATICS

In this paper we consider compulsory education in Norway. We use the present Norwegian terminology, *The Basic School*, to denote the type of school. Norway has a 9-year basic school. It is divided into 2 stages: primary school (Grades 1–6) and lower secondary school (Grades 7–9). We look mainly into formal assessment of students.

Historical Outline

Before 1938, a wide variety of assessment systems were in use in different regions. A widespread "inflation" in grades was observed. Following the national curriculum plans in 1938, a national system of evaluation was introduced. There were guidelines published for school-based assessment as well as national external examinations. For large groups of students, a norm-referenced grading scale was introduced with the percentages 4, 24, 44, 24, 4 for the hierarchy of grades.

In the mid-1950s, Norway started a process that extended compulsory education from seven to nine years. In this process, revised assessment guidelines were published in 1964. The guidelines have undergone several minor revisions. Nine years of compulsory education was established by a school law in 1969. The curriculum plan was completed in 1974; it is referred to as *M74* (Kirke- og undervisningsdepartementet, 1974).

In the process of extending compulsory education, a committee was formed in 1972 to consider assessment in schools. The two documents from the committee appeared in 1974 and 1978 (Evalueringsutvalget, 1974 & 1978). In the last of these papers the committee took a radical position concerning assessment:

> 'The majority of the committee proposes that there should be no formal assessment in lower secondary education [i.e. grades 7–9]' (Evalueringsutvalget, 1978, p. 33).

Formal assessment should be interpreted as the use of a grading scale; *informal* assessment should be interpreted as teachers' opinions expressed

M. Niss (ed.), Cases of Assessment in Mathematics Education, 107–118.

in written or oral fashion. This was a too radical change for the majority, and the proposal was met with vigorous opposition. For various reasons, this has not been an official policy since 1978. Norway now seems to be in a situation where questions about formal assessment are not being discussed.

The curriculum plan (*M87*) was revised in 1987 (Kirke- og undervisnings-departementet, 1987). Assessment was not discussed directly in the revised plan, and so far, there has been no attempt to perform a thorough revision of the assessment guidelines.

The Present Situation

The 1938 system is still in use today, with some modifications. There are no formal grades in primary school. The teachers give oral or written reports on performance to students and parents. In lower secondary school, the formal assessment is mainly school based. At the end of Grade 9 there is a written external exam (final), common for the whole country, and a possible oral examination.

The students have a written final exam in one of the subjects, Norwegian, English or mathematics. The subject is determined by a draw and is announced two weeks in advance of the test. Oral examinations have become more common in later years; now more than half of the students have an oral exam in one of the school subjects.

Both the school-based and the examination grades are reported at the end of Grade 9. A student might, therefore, get as many as three grades in mathematics. For further selection in the educational system, the average of the grades in each subject is computed.

Trends

Two observed trends in the present situation are worth noting. First, assessment subcultures are developing in primary school. Teachers have developed their own "formalized" way of reporting performance. Second, there are reports on the grade inflation in lower secondary school. Hence, the situation now is somewhat similar to the situation before 1938.

2. FINAL EXAMS IN MATHEMATICS

From 1984 to 1989 there have been two versions of the final exams in mathematics. One version permits the students to use a calculator for part of the exam and one version does not allow calculators. It is usually the school, or in some cases even the single teacher, that chooses the version that will be used. The choice depends on whether students have been using calculators in their coursework.

Both versions consist of two parts. One concerns basic skills (without a calculator) and the second consists of more extensive problems. In the second, the student is supposed to present a written solution, giving details of his method, whereas in the first part, the answer or computations are to be written on the exam sheet. Space is provided for this purpose. For the calculator version, a calculator is available for the second part. The two parts of the exam are given out at the same time. The student gets the calculator when the first part is handed in. The total time for both parts is 5 hours. The student is to self-determine when to hand in the first part. The system of assessment for the final exam is both central and regional. The guidelines for grading are made centrally a short time after the exam has been held. The persons responsible for administering the grading process in the regions meet in Oslo. Each person has graded as many exam papers as possible (usually from 100 to 150) before this meeting. During the meeting, the exam problems are discussed and each problem is given a certain "weight" (points). Then a recommendation is given on where the boundaries between the grades should be set The basis for this grading process is a normal distribution, with the sample of exam papers graded. The recommendations, along with a commentary on how to grade problems, are then mailed to the teachers doing the grading.

The assessment process is regional. Each paper is graded by two teachers in the region, who have to reach a common grade using a 5-point scale. The grading process leaves much to the teachers involved. There are no detailed grading schemes. The weight is on the overall impression of the student's paper. If both teachers have arrived at the same overall grade, performance on the individual problems is not discussed. The grades for a region are supposed to follow the normal distribution percentages of 4, 24, 44, 24, 4, and some adjustments may have to be made to meet this requirement. If that is the case, performance on individual problems will be discussed until an agreement is reached.

3. THE FINAL EXAM IN 1989

In 1989, three versions of the final exam were given: two versions of the 'traditional' exam (according to M74), with or without the use of a calculator, and one version following the revised curriculum plan, with a calculator. The three 1989 exams had several problems in common. In all three, the first part (basic skills) was identical. In the second part of the traditional exams, only two problems were different. The other problems were constructed so they did not favor the use of calculators. The exam following the revised curriculum plan had a different second part. Some of the problems were related to a common theme — the classroom. One problem in the second part was the same in all three exams.

Sample Problems

Traditional exams

We will concentrate on the second part of the exams. The problems are numbered from 11 to 17. Each problem may contain several questions denoted a, b, etc. The two sets are identical except for problems 13 and 16.

In *Problem 12a*, the formula for the volume (V) of a pyramid with a square base and a given height is presented:

$$V=\frac{a^2\cdot h}{3}$$

The problem is to express the height as a function of the side of the base and the volume.

Problem 14 is an equation to be solved and the result validated:

$$\frac{5+x}{2}+\frac{4(x+6)}{3}-(x+2)-3=\frac{5(x-3)}{3}-\frac{7x}{3}$$

Problem 15 is a traditional geometry problem with some construction and some computation. The numbers are chosen so as not to favor the students using calculators.

In the parallelogram $ABCD$, AB and CD are parallel. The angle ABC is 105 degrees, AB=10 cm, and BC=6 cm.

a) Draw or construct the parallelogram.
In this problem the task is to find the distance between AB and CD. To find this distance you will have to draw auxiliary lines. The point E is on CD such that the angle ABE is 60 degrees. The normal from E intersects AB in F.
b) Construct the line BE and the line EF.
c) Find the angles in the triangle BEF.
(The normal from C intersects BE in G.)
d) Find the angles in the triangle BCG and in the triangle CEG.
e) Find the length of CG.
f) Find the length of BE.
g) From the information you have, find the distance between AB and CD.

Revised curriculum exam

Here we find some new problem types:

Problem 13:

a, b, and c are three different positive integers less than 10 such that:

$$\frac{1}{a}+\frac{1}{b}+\frac{1}{c}=1$$

Find the three numbers a, b, and c.

Problem 18:

Given a scatter plot showing a linear relationship between the earth's mean temperature and the concentration of CO_2 in the atmosphere, the student is asked to:

a) Draw a line of best fit of the scatter plot.
b) Use the line to predict the earth's mean temperature at a CO_2 concentration of 1.15 ppm.
c) Use the line to predict CO_2 concentration when the earth's mean temperature is 8.5 degrees Celsius.
d) Choose from among 3 given equations the one that best describes the given points.
e) Explain why neither of the other two equations can be correct.

(The translation is taken from Romberg, Wilson & Chavarria (1990).)

Data on Students' Performance

In 1989, I was grading the papers from the exams in two regions. The task was to grade together with another teacher, about 150–200 student papers from 5–10 schools. Papers from all three versions of the exams were graded. In this process careful notes were taken on students' scores and how the problems were solved. In the two papers documented in this study (Gjone, 1990a & 1990b) an analysis showed that the student groups graded in traditional exams were representative of the region, with respect to grades. It should be noted that the only records of the written exams published by school authorities are the regional distributions of grades. The exam papers and grades of students are kept for some time, but the teachers' grading notes are not collected.

In the grading process a problem is graded right, wrong, or partially right. My grading was adjusted with the grading of the other teacher when differences in the overall grade were detected. Hence, there are uncertainties in the grades for individual problems. In this study, I have used two categories to mark problem solutions: right and not right.

4. ANALYSIS OF MATHEMATICS EXAMS

In two papers (Romberg, Wilson & Khaketla, 1989; Romberg, Wilson & Chavarria, 1990), a scheme for analyzing mathematical tests were presented. The rationale for this type of analysis is the influence of testing on teaching, documented in several studies (Romberg, Zarinnia & Williams, 1989). In Norway there also is a documented influence of testing (exams) on teaching. In a survey by *The Basic School Council* (Grunnskolerådet, 1990) the data in Table 1 were obtained.

In the first Romberg papers, several American tests were classified, and in the second some foreign tests (British) were considered in addition to American tests. The investigation was "undertaken to identify items, and

To a large extent	37%
To some extent	49%
To a small extent	9%
Not sure	1%

Table 1 *To which extent does the final exam influence your teaching (Grunnskolerådet, 1990, p.23)*

perhaps tests, that reflect the intent of the Standards' (Romberg, Wilson & Chavarria, 1990). Their reasons for including foreign tests were as follows:

"We felt it important to examine these because only in the United States have short answer, multiple-choice items been commonly used at each school level, while most other countries have not put such singular emphasis on arithmetic calculations." (p. 12)

They noted that "it was more difficult to classify British tests than American tests since the items could often be classified into more than one content, process or level category. In their analysis, each test item was classified to three dimensions: *content*, *process*, and *level of the response* required. The seven content areas were taken from NCTM *Standards* (Curriculum and Evaluation Standards for School Mathematics) (1989). The six process areas were: *communication*, *computation and estimation*, *connections*, *reasoning*, *problem solving*, and *patterns and functions*, also found in the Standards. The level of response was either *procedure* or *concept*.

We have analyzed Norwegian exams using the method developed by Romberg and his colleagues. There are several reasons for applying this method to Norwegian tests. The classification scheme can function as a tool for analyzing changes in exams over time, especially in the present situation, when Norway is in the process of implementing a new curriculum. The method can also be used to analyze differences in tests between different levels in the school system. It is also interesting to compare Norwegian tests (exams) with the tests of other countries.

The Model

We have used the same three dimensions — content, process, and level of response — as mentioned above. Our categories are only slightly different from the categories used by Romberg and his colleagues. It has been necessary to include more content categories, to more closely reflect the content areas in the basic Norwegian school curriculum[2].

In the content dimension, we used these categories: (an asterix * marks categories that are similar to those used by Romberg and his colleagues).

Numbers and Number Relations (*)
Measurement (*)
Percent
Geometry (*)
Statistics (*)
Economics
Algebra and Functions (*)

Our process areas are basically the areas used by Romberg and colleagues, but reformulated to correspond more closely to the Norwegian curriculum plan, Grades 1–9.

Communication
Computation and Estimation
Reasoning
Exploration of Patterns
Connections and Modeling
Problem Solving

Because of the nature of the problems posed in the Norwegian exams, it has been difficult to classify each item into one single category. An attempt, however, has been made to use a single category classification to be consistent with the criteria in Romberg, Wilson & Khaketla (1989). The classification has not been tested by others, it has revealed some characteristics of the various versions and students' performance.

Applications to Norwegian Exams

In this study, we considered the three versions of the final exam given in 1989. The exam followed the new curriculum plan (M87) and contained some new problem types. We asked how this exam fits the "profile" specified by Romberg and his colleagues. We also used his classification scheme as a basis for the study of student performance.

Our classification matrix is basically the same as that found in Romberg, Wilson & Khaketla (1989), with some added features. In the Norwegian exam, each problem/test item carries a 'weight' that reflects the difficulty and amount of work involved in solving the problem. The tests have been classified using the percentages representing these weights.

In the classification process, a certain problem is put into one of the categories of each of the dimensions. Let us illustrate the process with an example.

The first problem in the exam was to compute the sum:

312+56+1030

This problem is classified as follows:

In the content dimension the problem is classified as numbers and relations, in the process dimension as computation and estimation, and in the level dimension as procedure. As mentioned above, more complex problems pose difficulties in this classification process. Many problems on the Norwegian exam fall into several categories. Each problem has been classified into what is seen as a "main" category, e.g., if a percentage is found, percent is used, not number and number relations, even if numerical computations have been used.

After all the problems in the exam have been classified, the totals (weight) for each category in the content and process dimension are computed for each level (concept/procedure). As an example, it might be found that computation and estimation sum to 8 points (out of a total of 50 points) in the process dimension (and concept level). Hence in this dimension (and level), computation and estimation, is 16 percent.

Developments in the Final Exams

The classification of the three versions of the 1989 exams are summarized in Table 2.

The most notable feature of the comparison is that the exam given according to the revised curriculum shows a definite shift towards more weight on concept in the *level of response*. This is a dimension which had not been considered explicitly in the curriculum revision, and it is difficult to interpret this change.

Concerning the *content* categories, we find an overall increase in the numbers category and a decrease in economy. In algebra and functions, more stress is on concept and less on process. The traditional equations have been reduced in volume and difficulty; this corresponds with the guidelines in the revised curriculum. In geometry we find a shift towards balance between the concept and procedure levels of response.

There are also some interesting changes in the *process* categories. The "exploration of patterns" category has received increased weight, and so has the connections and modeling category. It is to be noted that there is also an overall reduction in the computation and estimation category. These changes are seen to correspond with the intended new curriculum. There is, however, one notable exception. Even if problem solving received more attention in the revised curriculum plan, we see that the exam contains less of what we call problem solving exercises. One of the reasons for this shift is the new form of geometry problems — with more weight on carrying out procedures and computation. The traditional geometry problems contained ruler and compass constructions which was classified as problem solving.

Table 2 *Exam results*

Test	Content							Process						Level	
	NNR	MEA	PRC	GEO	STA	ECO	ALG	COM	C&E	REA	EXP	C&M	PS		
M74	4	12	16	44	12	0	8	8	16	16	0	24	36	CONC	41
(NC)	17	17	8	8	0	17	33	0	69	0	8	6	17	PROC	59
M74	4	17	13	46	13	0	8	4	17	17	0	25	38	CONC	39
(C)	16	16	8	8	0	19	32	0	68	0	8	5	19	PROC	61
M87	16	10	10	26	16	3	19	6	13	10	23	35	13	CONC	51
	27	20	3	27	0	7	17	0	67	7	0	7	20	PROC	49

Content categories:
NNR Numbers and number relations
MEA Measurement
PRC Percent
GEO Geometry
STA Statistics
ECO Economy
ALG Algebra and Functions

Process categories:
COM Communications
C&E Computation and Est.
REA Reasoning
EXP Exploration of Patterns
C&M Connections and Modeling
PS Problem solving

M74-NC Traditional (No calculator)
M74-C Traditional (With calculator)
M87 Revised curriculum (1987)

CONC Concept
PROC Procedure

Note. The percentages are to be interpreted as follows (example) For M87, Measurement, we have 10% at Concept level and 20% at Procedure level. This means MEAsurement-CONCept is close to 5%, and MEAsurement-PROCedure is close to 10% of the total Moreover, MEAsurement makes up 15% of the content of this exam.

One must be careful when drawing conclusions from these few data. To investigate development, it is necessary to look at a wider range of final exams both before and after 1989. However, one should not underestimate the "signals" that this one exam might give teachers, since all three versions are easily compared.

Student Performance

To give an idea of the results obtained, we present a table of student scores on the problems presented above:

Test	Sex									
		12a	14	15a	15b	15c	15d	15e	15f	15g
M74 NC	B	23	14	63	65	45	32	28	18	26
	G	26	9	58	60	47	40	31	16	17
M74 C	B	44	32	82	85	66	45	41	33	33
	G	43	47	70	87	67	57	37	20	23
		13	18a	18b	18c	18d	18e			
M87	B	23	73	10	13	47	20			
	G	21	65	12	15	53	21			

Note The numbers are percentages of students having a correct answer (B:boy, G:girl). One observation that can be made is that on the M74 version, students with calculators performed better than students without calculators This can be seen on other problems as well.

Table 3 *Student scores*

There are many ways to combine the information obtained by relating student performance to the categories used for items. However, because of the comparatively small number of problems in each category, specific conclusions can not be inferred.

If we look at the five problems with the lowest scores (all below 15 percent) in the revised curriculum exam, four are classified as concept level. In the process area they belong to the categories of reasoning, exploration of patterns, connections and modeling, and problem solving. If we consider some of these observations in conjunction with the development of the exams, we see that the same process categories also gave increased weight in the revised curriculum exam. We exclude problem solving in this discussion, since the reduction of this category is due to one type of problem.

We should be cautious about inferring a development from an analysis of this first exam after one year of the new curriculum. However, if students perform more poorly on a large part of the informal exam we may have a shift in the meaning of the grades in the external exam. A student may get a good grade with a comparatively weak performance with the norm-based scale. This, in turn, may have an effect on the observed inflation of internal grades. Conversely, if we look at the three problems with the highest scores in the second part of the same exam (all above 80 percent), they belong to communication, and computation and estimation. This result is hardly surprising, since the students have a comparatively large amount of training on these problem types.

There has been some unrest among teachers concerning new problem types in recent years. Many arguments have been put forward in favor of the traditional problem type. It is therefore with some surprise we find the following: The problem with the lowest score (about 29 percent) in part 1 (counting all versions of exams) was to simplify the expression:

$$x\cdot(2x-1)+\frac{x}{4}\cdot 8+x(x-1)$$

This problem was classified as computation, algebra, process. This result is difficult to explain, since this is a standard problem in exams and textbooks.

5. CONCLUDING REMARKS

The content and development of mathematics tests have not been studied systematically in Norway. In the Norwegian system, where much of the assessment process is informal and subjective, it is important to be able to "measure" profiles of tests as a basis for further development. The use of the classification scheme, as developed by Romberg and his colleagues, shows that there are some possible developments in Norwegian mathematics tests that need to be analyzed further.

NOTE

The research reported in this paper was supported by the Program for Research on Education under the Norwegian Research Council for Science and the Humanities and the Regional Board of Colleges in Oslo/Akershus.

1 These are not documented trends, but based on unsystematic observations and articles in newspapers and teacher journals.

2 For analyzing differences between tests at different levels of a school system, as well as differences between countries, some common content categories should be agreed upon

REFERENCES

Evalueringsutvalget for Skoleverket: 1974, *Karakterer, Eksamener, Kompetanse m.v. i Skoleverket* NOU 1974:41 (in Norwegian), Universitetsforlaget, Oslo.

Evalueringsutvalget for Skoleverket: 1978, *Vurdering, Kompetanse og Inntak i Skoleverket* NOU 1978:2 (in Norwegian), Universitetsforlaget, Oslo.

Gjone, G.: 1990a, 'Resultater fra avgangsprøva i matematikk 1988', in *Elevprestasjoner i Matematikk* (in Norwegian), Oslo College of Education, Oslo.

Gjone, G.: 1990b, 'Resultater fra avgangsprøva i matematikk 1989', in *Elevprestasjoner i Matematik II* (in Norwegian), Oslo College of Education, Oslo.

Grunnskolerådet: 1990, *Rapport. Seminar om Styrking av Matematikkundervisningen etter M87* (in Norwegian), Oslo.

Kirke- og Undervisningsdepartementet: 1984, *Mønsterplan for Grunnskolen* M84 (in Norwegian), Aschehoug, Oslo.

Kirke- og Undervisningsdepartementet: 1987, *Mønsterplan for Grunnskolen* M87 (in Norwegian), Aschehoug, Oslo.

National Council of Teachers of Mathematics (NCTM): 1989, *Curriculum and Evaluation Standards for School Mathematics*, Reston, VA.

Romberg, T.A., Wilson, L. & Chavarria, S.: 1990, *An Examination of State and Foreign Tests*, National Center for Research in Mathematical Sciences Education, Madison, WI.

Romberg, T.A., Wilson, L. & Khatela, M.: 1989, *An Examination of Six Standard Mathematics Tests for Grade Eight*, National Center for Research in Mathematical Sciences Education, Madison, WI.

Romberg, T.A., Zarinnia, E.A. & Williams, S.: 1989, *The Influence of Mandated Testing on Mathematics Instruction: Grade Eight Teachers' Perception*, National Center for Research in Mathematical Sciences Education, Madison, WI.

Gunnar Gjone
University of Oslo,
(formerly at Oslo College of Education)
Norway

HANS NYGAARD JENSEN

ASSESSMENT OF PRIMARY AND LOWER SECONDARY MATHEMATICS IN DENMARK

1. INTRODUCTION: MATHEMATICS IN PRIMARY AND LOWER SECONDARY SCHOOL

The primary and lower secondary school in Denmark, called the Danish *Folkeskole*, is a comprehensive school enabling children to remain in the same pupil group from the 1st to the 9th or 10th Grade, as they progress automatically from one class to another, irrespective of yearly attainment. The aim of the school is to give pupils the possibility of acquiring knowledge, skills, working methods, and ways of expressing themselves that will contribute to their all-round development as individuals.

Mathematics instruction takes place during every Grade of the school, and the main areas of the subject, numbers and algebra, geometry, and statistics and probability, are taught at every level. The aims for instruction are described in general terms and include references to concepts, skills, and attitudes, as well as to content. Each local municipality has to adopt the aims laid down by the Ministry of Education for a subject but is free to develop its own guidelines, which then becomes mandatory in its school. The vast majority of the municipalities adopt the Ministry guidelines for all or almost all subjects.

In the 8th to 10th Grade, the pupils will have to choose between mathematics instruction given in *a basic course* or in *an advanced course*. It is also possible for schools, in cooperation with the parents, to offer *unstreamed courses* in mathematics. During the years, this option has been more common so that, today, almost 90 percent of all pupils are following either an advanced course (33 percent) or an unstreamed course (55 percent). It is a movement towards "mathematics for the mass", so to speak.

2. ASSESSMENT IN MATHEMATICS

Internal assessment may be done by the teacher in an informal way during the school course. The parents will be informed formally about their children's progress at least twice a year. The general mode of reporting is oral, but after the 8th Grade, the assessment has to include a written report.

M. Niss (ed.), Cases of Assessment in Mathematics Education, 119–127.

The external assessment system in mathematics consists of *The School Leaving Examination* and *The Advanced School Leaving Examination*. The School Leaving Examination may be taken after both the 9th and the 10th Grades. All pupils present themselves for these examinations, no matter whether they have been taught in a basic course, an advanced course, or an unstreamed course. The examination in mathematics can only be taken in a written form. It consists of a one-hour test in basic skills with answers written on the examination paper by the pupil, and a four-hour written paper containing problems mainly of a practical and applicational nature.

The Advanced School Leaving Examination can be taken only after 10th Grade, and only by pupils who have taken advanced or unstreamed courses. The examination has oral and written parts. The four-hour written examinations have the same characteristics as the School Leaving Examinations, but have more and higher level problems. The pupils can proceed in the educational system — to the different kinds of education for youth — both after the 9th and after the 10th Grade. So, the Advanced School Leaving Examination is not required, either in principle or in practice, to enter, for example, the upper secondary school, the *Gymnasium*. After the 9th Grade, 55 percent of pupils will continue to the 10th Grade, while 21 percent will proceed to upper secondary school, and 15 percent will join *Basic Vocational Training*. After the 10th Grade, 23 percent of pupils will continue to the upper secondary school or to other courses at the same level.

Examinations are not compulsory. The pupil has the right to decide whether or not to sit for them, after consultations with the school and his parents. It is, however, rare for a pupil to decide not to sit for an examination: Only about 2 percent do not participate after the 9th Grade and 1.5 percent after the 10th Grade.

Each examination subject is assessed on its own merit so the results cannot be summed to give an average mark. No one can "flunk out" of the public education system because of low examinations, although pupils obviously can obtain insufficient marks to continue to upper secondary or tertiary education.

The assessment scale for achievement is divided into three main groups good, average, and weak, and contains 10 different marks in all. The written examinations are marked relatively to the national average, according to centrally-determined criteria by centrally-appointed external examiners.

In addition to the final examination, the pupil will be assessed for the year's work, and although the same scale is used, he is now judged according to the class average. It can be a problem, especially for the pupil from an advanced course taken after the 9th Grade, to be judged in the same subject but on such a different basis.

Oral examinations occurring after the 10th Grade are set locally. They are carried out locally, but, since from 1990, to some extent, with the participation of external examiners.

3. CONTENTS AND APPEARANCE OF THE WRITTEN EXAMINATION PAPER

The written examination paper after the 9th and the 10th Grades has gradually changed in appearance and content during the last 10 years. The written paper now deals with an overall topic, which is elaborated in different ways in the specific parts of the examination.

Examples of such *thematic examinations* are:

- Calculations and problem solving in connection with a craft (e.g., carpenter, farmer, nurse, fireman), its education, economic conditions, materials, and statistical data.
- Calculations and problem solving in connection with having a baby (e.g., the growth and weight of the baby, salary problems caused by maternity leave, etc.).
- Calculations and problem solving in connection with a birthday party (e.g., shopping, baking a cake, table arrangement).

And so on.

The following (Figure 1) example comes from the 9th Grade examination of May-June, 1989. The theme in this paper deals with the sport of table tennis. The problems and concerns are with buying equipment, the area of the table or the playing field, the speed of the tennis ball, determining the number of games by combinatorics, and a description using functions of different models for collecting money to support the tennis crew. Finally, the last page deals with packing the balls, and includes an open-ended question about comparisons of different types of boxes.

The following three pages, **Figure 1**, contain the *School Leaving Examination, 'problem arithmetic', 9th Grade, May-June, 1989, 4 hours*. An English translation is sebsequently presented, **Figure 2**.

1. KØB AF UDSTYR TIL BORDTENNIS

En bordtennisklub køber følgende udstyr:

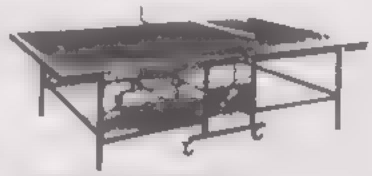
15 æsker bolde
12 trøjer
2 bordtennisnet
2 bordtennisborde

- Hvor meget koster udstyret i alt?
- Hvor meget skal klubben betale, hvis den får 15% rabat?

Æske med 3 bolde 33 kr.


Net 410 kr.


Bordtennistrøje 185 kr.


Bordtennisbord 5 210 kr.

2. SPILLEPLADSEN

Til en bordtenniskamp skal spillepladsen være 12 m × 6 m.

- Beregn spillepladsens areal.

Gulvet i en sportshal er 40 m × 20 m.

- Hvor mange spillepladser kan der højst være i hallen?

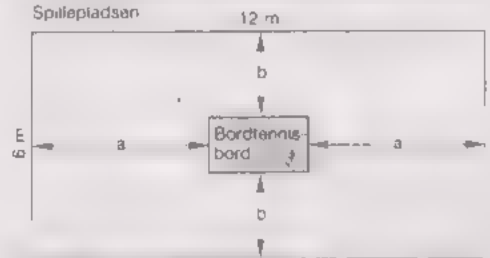


fig. 1

På figur 1 er bordtennisbordet anbragt på midten af spillepladsen.

- Beregn afstandene a og b i virkeligheden.


Bordtennis-
bordet

3. VERDENS HURTIGSTE SPIL

En bordtennisbold kan under spillet opnå en fart på 170 km i timen.
Denne formel kan bruges til at beregne farten:

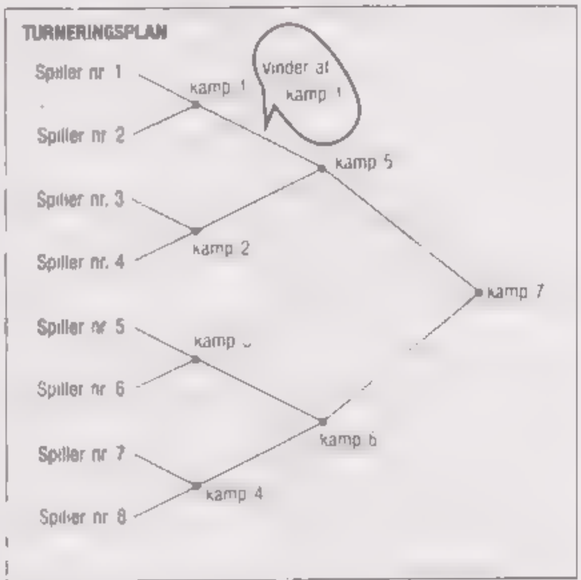
$V = 3{,}6 \cdot s : t$	**V** er farten, målt i km pr. time **s** er den strækning, bolden har bevæget sig, målt i meter **t** er tiden, målt i sekunder

- Beregn farten V, når s = 4,5 m og t = 0,1 sekund.
- Hvor mange meter har en bold bevæget sig på 0,25 sekunder, når farten er 90 km pr. time?
- Beregn tiden t, når s = 6 m og V = 144 km pr. time.

4. CUP-TURNERING

I en **cup-turnering** er det kun vinderen af hver kamp, der går videre.

I en cup-turnering deltager 8 spillere.
På figuren kan du se planen for turneringen.



- Hvor mange kampe har vinderen af turneringen spillet i alt?
- Hvor mange kampe er der i alt i en cup-turnering, hvor der er 16 spillere?

Her er vist en turneringsplan for 3 spillere.

- Tegn en turneringsplan for 6 spillere.
- Udfyld et skema som dette.

[illegible]	[illegible]	[illegible]	5	6	7	8
[illegible]						
[illegible]						

- Angiv det samlede antal kampe i en cup-turnering med x spillere.

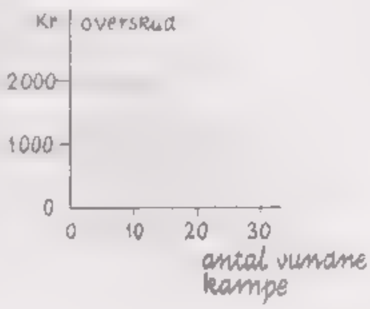
5. INDSAMLING TIL EN REJSE

[illegible] penge til en rejse. De opfordrer derfor venner og bekendte til at støtte holdet [illegible] støtte holdet. Hver person skal give [illegible] kr. når der vindes en kamp i holdturneringen. Der er faste udgifter på 400 kr. i forbindelse med støtteordningen.

- Hvor stort et beløb giver støtteordningen i overskud, hvis holdet vinder 52 kampe?

Det samlede overskud kan udtrykkes ved ligningen

MODEL I	$y = 100 \cdot x - 400$	y er overskuddet angivet i kr. x er antal vundne kampe



- Tegn grafen for denne ligning i et koordinatsystem som det viste.

Hvis holdet i stedet får støtte fra 125 personer og har faste udgifter på 1 000 kr., kan overskuddet udtrykkes ved denne ligning.

MODEL II	$y = 125 \cdot x - 1000$	y er overskuddet angivet i kr. x er antal vundne kampe

- Tegn grafen for denne ligning i samme koordinatsystem.
- Hvor mange kampe skal der vindes efter hver af de to modeller, hvis overskuddet skal være på 10 000 kr.?
- Hvor mange kampe skal der vindes, for at overskuddet bliver det samme efter de to modeller?

6. ÆSKER TIL BOLDE

Bordtennisbolde kan købes i æsker med 3 bolde – se figur 1.

- Beregn rumfanget af æsken.

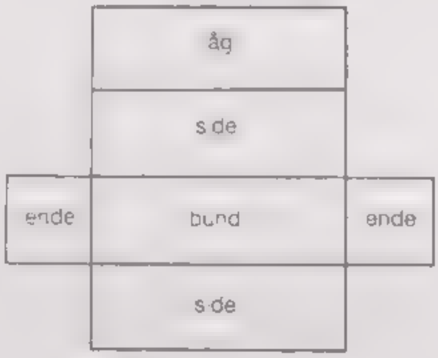
Rumfanget af en bold er 28,7 cm^3.

- Hvor mange cm^3 luft er der uden om de 3 bolde i æsken?
- Hvor mange procent udgør denne luft af æskens rumfang?

figur 1

Udfoldet kan papæsken se ud som vist på skitsen – se figur 2.

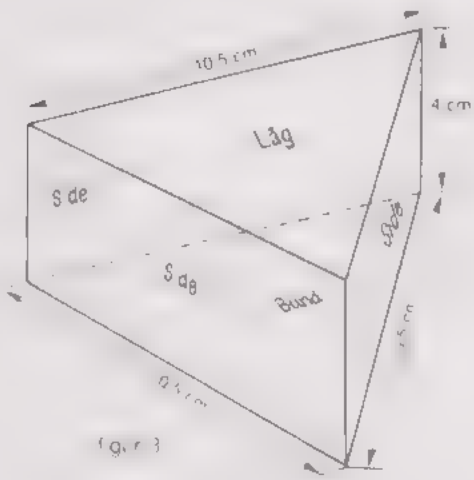
- Beregn arealet af dette stykke pap.



figur 2

På figur 3 er vist en anden æske, hvori der også kan være 3 bolde

- Tegn en nøjagtig figur af den udfoldede æske på figur 3 i naturlig størrelse.
- Foretag en sammenligning af de 2 æsker, idet du blandt andet kan foretage beregninger som ovenfor
- Vis på din tegning med cirkler, hvordan de 3 bolde er placeret i æsken



figur 3

Problem arithmetic: the School Leaving Examination, May-June 1989, 4 hours

1. Buying table tennis equipment
A table tennis club buys
15 boxes of balls
12 t-shirts
2 nets
2 tables
(at the prices shown in the figure).
- How much is the equipment in total?
- How much does the club have to pay if it obtains a 15% deduction?

2. The playing field
A table tennis playing field is 12 m times 6 m.
- Compute the area of the playing field.

The floor in a sports arena is 40 m times 20 m.
- What is the maximum number of playing fields that can be placed in the arena?

In Figure 1, the table is placed in the center of the playing field.
- Calculate the real distances a and b.

3. The fastest game in the world
A table tennis ball may obtain a speed of 170 km/h. The following formula can be used to calculate the speed

$$V = 3{,}6 \cdot s \cdot t$$

where V is the speed (km/h), s is the distance travelled by the ball (m), and t the time (seconds).
- Compute the speed V, if s=4.5 m, and t=0.1 secs
- How many meters has a ball moved in 0.28 secs, if the speed is 90 km/h?
- Compute the time t, if s=6 m, and V=144 km/h.

4. A cup tournament (only the winner of each match stays in the tournament)
We look at a cup tournament with eight players. The tournament schedule is shown in the figure.
- How many matches has the winner of the tournament played altogether?
- What is the total number of matches in a cup tournament with sixteen players?

This is a schedule for a three-player tournament.
- Draw a schedule for a six-player tournament.
- Complete the incomplete table shown.
- Indicate the total number of matches in a cup tournament with x players.

5. Raising funds for a trip
A table tennis team needs money for a trip. They invite friends to sponsor the team. They succeed in finding a 100 sponsors. Each individual promises to give 1 krone for a match won in the team tournament. The fixed costs amount to 400 kroner.
- What is the size of the net profit if the team wins 52 matches?

The net profit can be calculated according to

Model I: $y = 100x\text{-}400$

(y = the net profit (kroner), x = the number of matches won).

- Draw the graph corresponding to the equation in a coordinate system.

If instead the team is sponsored by 125 persons and has fixed expenditures of 1,000 kroner, the net profit can be expressed by

Model II: $y = 125x\text{-}1000$

(x and y as before).

- Draw the graph corresponding to the equation in the same coordinate system as above.
- For the net profit to be 10,000 kroner, how many matches have to be won in each of the two models?
- How many matches have to be won for the net profit to be the same in both models?

6. Boxes for balls

Table tennis balls are bought in boxes of three balls (see figure 1).

- Calculate the volume of a box.

The volume of one ball is 28.7 cm^3.

- How much air (in cm^3) is surrounding the three balls in the box?
- How big a percentage of the box volume does this amount of air represent?

If the box is unfolded it can look as in figure 2.

- Calculate the area of this piece of cardboard.

Figure 3 shows a different type of box that can also hold three balls.

- Draw an accurate real-size picture of the box in unfolded shape.
- Compare the two boxes, e.g. by performing calculations similar to those mentioned above.
- Show in your own drawing, by means of circles, where the three balls are placed in the box.

Figure 2 *Translation of Figure 1*

The 10th Grade examination will, as mentioned, have greater length (normally about 5 pages) and high-level problems, including some more abstract problems from the theory of functions (quadratic and exponential functions), statistics, and probability.

4. STRENGTHENING THE LINGUISTIC COMMUNICATION IN MATHEMATICS EDUCATION

A report from the Danish Ministry of Education (1990), concerning content and quality in mathematics in the Danish educational system, recently pointed out that linguistic communication — oral as well as written — should be strengthened in several parts of the system. As regards the Danish

Folkeskole, it recommended that the 9th and 10th Grade assessments include oral examination. Looking at the experiences with the 10th Grade oral examination, it recommended that it not be an imitation of the written examination, but rather the pupil's presentations of experiences and backgrounds derived from a subject they develop individually in connection with their daily work.

REFERENCES

Danish Ministry of Education and Research: 1990, *Mathematics: Quality in education and teaching* (in Danish), Copenhagen.

Hans Nygaard Jensen
N. Zahle's College of Teacher Education,
Copenhagen,
Denmark

KIRSTEN HERMANN & BENT HIRSBERG

ASSESSMENT IN UPPER SECONDARY MATHEMATICS IN DENMARK

1. INTRODUCTION: THE DANISH SCHOOL SYSTEM

The Danish school system is divided into two sub-systems that function as two separate systems with regard to governance and teacher recruitment. Pupils between 7 and 16 years of age are taught in the *Folkeskole* (primary and lower secondary education), a 9 or 10-year compulsory school, whereas the non-vocational teaching of the 16–19 year olds (upper secondary education) is provided by the three year high school/grammar school called the *Gymnasium* that admits just under 30 percent of a cohort (approximately 20,000 students). Instruction in the Gymnasium provides both further education and general education. The teachers hold university degrees, usually in two subjects, the levels of which correspond to a master's degree.

2. MATHEMATICS INSTRUCTION IN THE GYMNASIUM

The Gymnasium has a linguistic and a mathematical stream. Mathematics is not compulsory for the students in the linguistic stream, although elements of the subject form part of a science course. Therefore, the following will deal exclusively with mathematics instruction in the mathematical stream of the Gymnasium. Mathematics in this stream is taught at two levels, called A and B. B-level mathematics is achieved at the end of the first two years with 5 lessons (45 minutes each) a week, and A-level is achieved at the end of the third year with a further 5 lessons a week. A-level is attended by 80 percent of the students from the B-level.

Mathematics instruction comprises: pure mathematical topics and three so-called *aspects* of mathematics, historical, modeling, and structural.

Pure mathematical topics

At B-level: number theory, geometry (including trigonometry), functions, differential calculus, statistics, and probability; At A-level: integral calculus, differential equations, vector theory, geometry in two and three dimensions, and computer-oriented mathematics.

M. Niss (ed.), Cases of Assessment in Mathematics Education, 129–137.

Three Aspects
At B-level as well as A-level: The *historical* aspect aims at familiarizing the students with element of the history of mathematics and mathematics in cultural and social contexts; The *models and modeling* aspect aims at making the student familiar with the building of mathematical models as representations of reality; they are given an idea of the potentials and limitations in the application of mathematical models. In addition, the instruction should enable them to carry out a not-too-complex modeling process; The *internal structure of mathematics* aspect aims at providing students with an understanding of the modes of thought and the methods characteristic of mathematics and their contributions to the development and structuring of mathematical topics.

3. ASSESSMENT IN MATHEMATICS

The students in mathematics are assessed internally several times in the course of the three years. The teacher and the students agree between them how the internal assessment it to be carried out For all students, the final examination at B-level comprises a written examination paper (4 hours) prepared by central authorities (the Ministry of Education), and for students not continuing to A-level, an oral examination (approximately 25 minutes). Not all students in the country are given an oral examination every year. The Ministry of Education selects a number of students/classes to be examined at the very end of the school year. The final examination at A-level comprises a written examination paper (4 hours) and — again for a sample of classes — an oral examination (approximately 30 minutes).

The Written Examination Paper

The written examination paper contains purely mathematical problems as well as problems involving applications and simple modeling. Most problems are compulsory, but there are also a few optional problems. At B-level approximately 25 percent of the problems are simple, aiming at differentiating between different categories of less able students. The remaining problems are more complex in nature.

A couple of problems of more complex nature from the *1990 B-level* written examination follow:

Example 1

> A factory discharges a phosphorous compound into its waste water During the period 1986–87, the waste water is examined, and it appears that the daily discharge of the phosphorous compound is normally distributed with a mean of 0.53 kg and a standard deviation of 0.20 kg.

Q. Determine the probability of discharging more than 0.70 kg of phosphorous compound during an arbitrary day in the period.

Q. Estimate the amount of daily discharge of phosphorous compound on the 10 percent of days when the discharge is at its lowest.

Efforts to clean the waste water of the factory are increased. Thus at the end of 1988 the daily discharge of the phosphorous compound was normally distributed with a mean of 0.41 kg, and the probability of discharging more than 0.70 kg during an arbitrary day decreased to 2 percent.

Q. Determine the standard deviation of the daily discharge of phosphorous compound at the end of 1988.

Example 2

Q. Examine the function f with respect to domain, zero-points, signs, and monotonicity, with f defined by

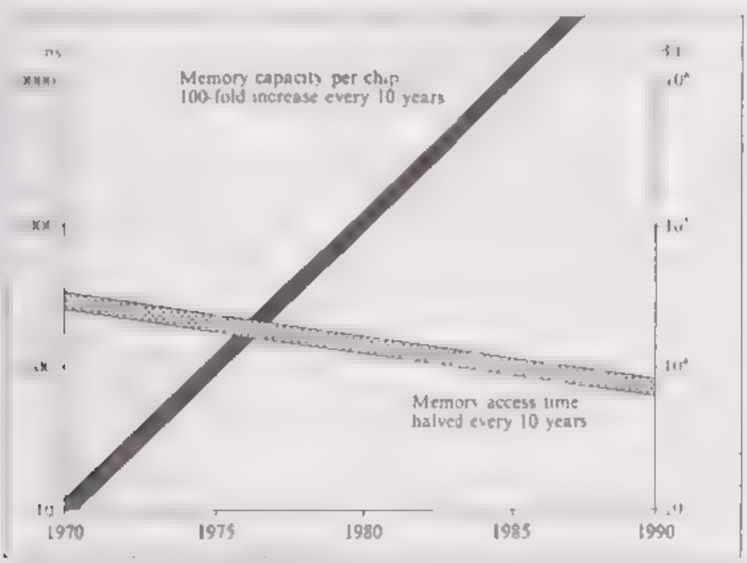
$$f(x)=\frac{x^2+6x}{x-2}.$$

Q. Determine each asymptote of the graph.

Q. Draw the graph of f, and determine the set of values of f.

Example 3

This excerpt is from "How a computer is developed and produced" published by Siemens A/G in 1989.



Q. By how many percent is the memory capacity per chip increased every year?

Q. Determine the double-value period of the memory capacity per chip.

Q. Find an expression for a function that describes how the memory access time has developed since 1970 at which time it was 300 ns.

Q. In which year will the access time be 40 ns. if the illustrated development goes on unchanged?

As will appear from the above examples of applications, the students are not required to mathematize problems themselves, except in very simple cases, i.e., giving arguments for linear, exponential, and power growth or — as in the example above — to find formulas for such functions.

Each student's answers are assessed by two officially appointed examiners. Their assessment results in examination marks based on a 10-step scale.

The requirements of the students' answers are the following:

- The correctness of methods and calculations applied by students
- The clarity of the students' modes of thought as seen from the answers

A good answer therefore comprises:

- Clear and well-arranged calculations
- Carefully executed drawings (i.e., diagrams, graphs, and geometrical figures)
- An explanatory text clarifying the way the problem is solved, with an emphasis on the (sub-) conclusions reached

The Oral Examination

The topics for the oral examination are chosen with respect to their many-sidedness, degree of difficulty, and appropriateness for an oral examination, i.e., the students should be given the opportunity to show their understanding of mathematical concepts and modes of thought to a greater extent in the oral than expected in the written examination paper. At B-level, half the syllabus is chosen for oral examination and one ore more of the three aspects are included. At A-level two-thirds of the syllabus is chosen for examination.

All examinations are undertaken by the teacher and an officially appointed examiner. The external examiners are mostly teachers from other schools, and the assignment as an external examiner is a part of a teacher's job. The teacher produces the examination questions and examines the students. The external examiner has the role of listener, and is only allowed to put questions to the student through the teacher. The examination questions are chosen in such a way that the examination reflects the way the topics have been treated in everyday instruction. The oral examination

tests the students' abilities to explain essential parts of a mathematical topic and their general knowledge of a mathematical theme. As a consequence of these aims, the examination questions are divided into two parts — a headline giving the theme for the examination, and a more detailed description of a limited part of the theme, which the student is expected to explain unassisted.

At B-level the examination is divided into two parts, the latter being a conversation between the teacher and the student for the purpose of examining the student's general knowledge. At A-level the student is expected to show his general knowledge in a more unassisted way.

Some examples of examination questions follow:

Example 1 (B-level). Growth Models

- Expound on the exponential growth model, including formula and graph (the following data may be used as a basis).

Under favorable circumstances the bacteria escherichia coli makes a cell division every 20 minutes:

Time (min)	0	20	40	60	80	.
Number	1000	2000	4000	8000	16000	..

The hour wages (in Danish kroners) of female workers in Denmark were in the years 1963–70:

Year	1963	1964	1965	...
Hourly wage	5.97	6.61	7.43	..

At the oral examination it is expected that the student, in an unassisted way, will explain, e.g.,

- What sort of growth in the real world may be exponential and why?
- Why does $f(x)=ba^x$ describe constant growth in percentage?
- Why is $f(x)=ba^x$ equivalent to the graph of f being a straight line in a semi-logarithmic coordinate system?
- Why is the double-value period a meaningful concept?
- Why is

$$T_2 = \frac{\log 2}{\log a} \; ?$$

In the following discussion between the student and the teacher, other growth models are covered (in a general way); one or more of the following themes may be dealt with:

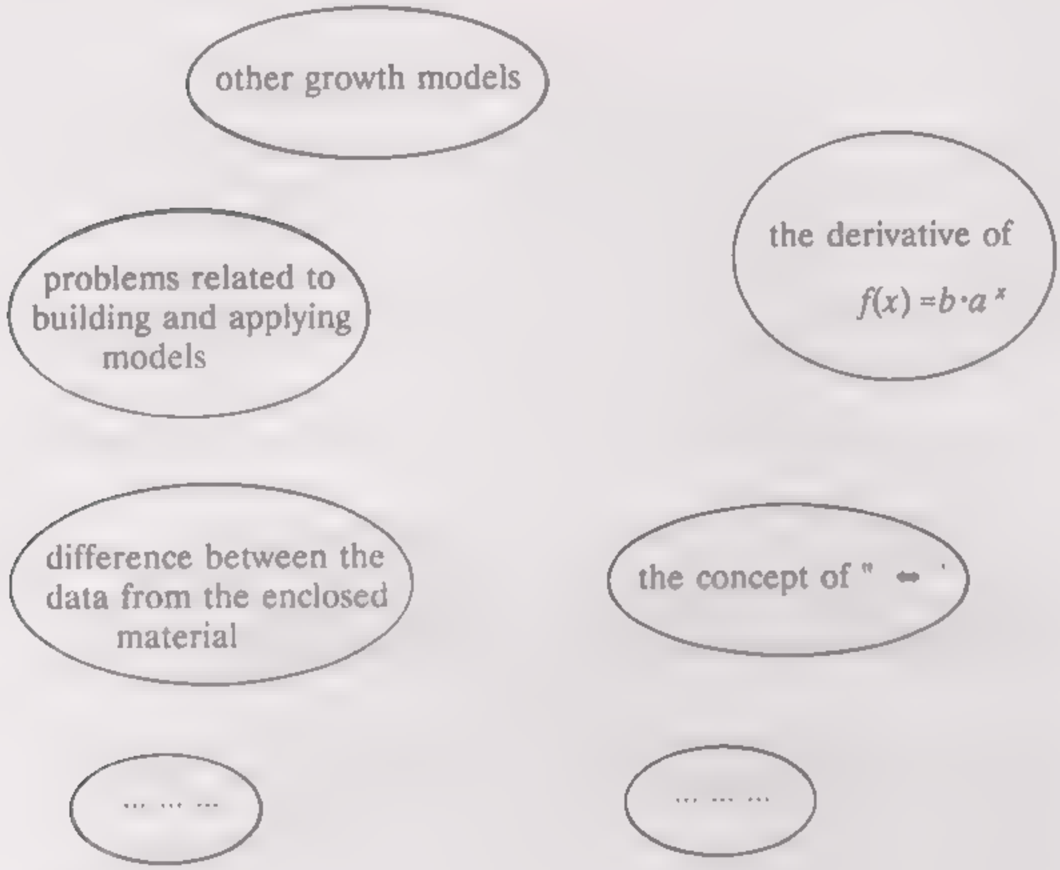


Figure 2

Other examples of questions are:

Example 2 (B-level). Equations of first and second degree

Q Expound on Descartes' Method in connection with the solution of the equation $zz = az + bb$.

Example 3 (B-level). Polynomials

Q. Expound on graphs and/or roots for polynomials of the second degree

Example 4 (A-level). Differential equations

Q. Expound on the logistic growth model. The enclosed material may be used as a basis. (This material is not included here).

The assessment of the students' abilities results in examination marks on the 10-step scale. The different forms of assessment and their relations to the major components of the curriculum are summarized in Tables 1 and 2.

Component \ Procedure		*Internal assessment*	
		Written	Oral
Main topics		Skills and understanding through: • exercises • problems	Explain • results • proofs • mathematical topics
Aspects	Historical aspect	E.g. essays (ex: "The golden section")	Small lectures e.g. "The Babylonian number system"
	Models and modeling	E.g. small project reports (ex: "Radioactive decay")	Small lectures e.g. "what is a mathematical model"
	Internal structure of mathematics	E.g. essays (ex: "The theorem of Pythagoras")	Explain e.g. "types of proofs"

Table 1 *Internal assessment*

4. A MAJOR WRITTEN PROJECT

In the third year, all students have to prepare a major written project in one of their subjects. Students' attending A-level mathematics may write their projects in mathematics. The students are rather free to choose the mathematical topic, and they often choose a topic related to one of the three aspects, but purely mathematical topics are also chosen.

Examples of titles of written projects are:

- The history of ... (e.g. the complex numbers)
- The babylonian number system
- Greek Mathematics: Geometric algebra
- Perspective drawings
- Mathematical models in decision making processes
- Mathematical models in epidemiology
- Chaos and fractals

<table>
<tr><td colspan="2" rowspan="2">Procedure / Component</td><td colspan="2">External assessment</td></tr>
<tr><td>Written</td><td>Oral</td></tr>
<tr><td colspan="2">Main topics</td><td>Skills in
• applying methods
• problem solving</td><td>Understanding of
• modes of thought
• mathematical themes</td></tr>
<tr><td rowspan="3">Aspects</td><td>Historical aspect</td><td></td><td>E.g greek mathematics</td></tr>
<tr><td>Models and modeling</td><td>• Applying known models
• Simple modeling</td><td>• Model-building process
• Application of mathematical models</td></tr>
<tr><td>Internal structure of mathematics</td><td>• "Explain that ..."</td><td>• concepts
• ideas
• proofs</td></tr>
</table>

Table 2 *External assessment*

The purpose of this project is to have students demonstrate their ability to learn about a mathematical theme and to put their acquired knowledge into written form in 10–15 pages. The project is assessed by the teacher and an external examiner, and their assessments also result in examination marks on the 10-step scale.

5. FINAL REMARKS

The Danish Ministry of Education recently has implemented the *Content and Quality Development Project*. In connection with this project, a report has been published on mathematics as a subject in the Danish Educational System (Danish Ministry of Education, 1990).

One of the points mentioned in the report is that linguistic communication — oral as well as written -should be playing a larger role in mathematics education. Therefore, it is very likely that in the future, oral examinations will be introduced in other parts of the school system where written examination papers are now the only form of examination.

REFERENCES

Siemens A/G 1989, *How a computer is developed and produced*, Siemens A/G, Denmark.

Danish Ministry of Education and Research. 1990, *Mathematics: Quality in education and teaching* (in Danish), Copenhagen.

Kirsten Hermann
Marselisborg Gymnasium,
Århus

Bent Hirsberg
Rosborg Gymnasium,
Vejle

both formerly at:
The Ministry of Education,
Copenhagen,
Denmark

WIM KLEIJNE & HENK SCHURING

ASSESSMENT OF EXAMINATIONS IN THE NETHERLANDS

1. INTRODUCTION

Secondary Education in The Netherlands is divided into four well-defined streams:

- Lower Vocational Education (Four years, 21 percent of the students),
- Intermediate General Education (Four years, 34 percent of the students),
- Higher General Education (Five years, 24 percent of the students),
- Pre-University Education (Six years, 20 percent of the students).

In 1985 a new program was introduced for the last two school years of pre-university education. It divided mathematics into two courses: *Mathematics A*, with an emphasis on applying mathematics in other subjects; and *Mathematics B*, emphasizing pure mathematics. The curriculum for Mathematics B is a variation development of the old program: analysis including calculus and differential equations, and geometry with a focus on 3-dimensional solids. The curriculum for Mathematics A differs fundamentally from that for Mathematics B: applied analysis including the derivative as a measure of change; applied algebra including matrices and linear programming; probability and statistics including hypothesis testing, informatics, and simple programming.

Mathematics A is designed for those who do not see mathematics as forming a substantial part of their future university studies, e.g. economics, psychology, etc. At the end of the pre-university course, students have to take a final examination in seven subjects. Mathematics is not compulsory, so students can take Mathematics A, Mathematics B, or both A and B, or no mathematics at all. In 1990, 59 percent of all pre-university students opted for Mathematics A, 47 percent for Mathematics B, and 19 percent for both A and B, leaving 13 percent of students opting for no mathematics.

Reform of the pre-university education curricula preceded that of the higher general education curricula. In this stream, students have to take a final examination in six subjects. From 1992 on, students can choose between two mathematics curricula: Mathematics A or Mathematics B. The examination syllabus for A consists of tables, graphs, formulas, discrete

M. Niss (ed.), Cases of Assessment in Mathematics Education, 139–154.

mathematics, statistics, and probability. The syllabus for B consists of applied analysis and geometry in three dimensions. A governmental working group currently is devising new mathematics programs, for all school streams for the age range 12–16, that add more emphasis on applied mathematics and 3-dimensional geometry.

Half of the assessment in the final year of secondary education is made up of teacher-made tests that are often written essay tests, although they sometimes include oral and sometimes individual pieces of work. As these tests differ from school to school and depend on the textbook used, it is not a simple matter to evaluate these tests.

We will restrict ourselves to the final examination papers that are valid for the whole country, and make up the other half of the final-year assessment. As the examination for mathematics includes open-ended questions, it is not easy to obtain data on the results. The Institute of Educational Measurement (CITO) asked every school to send us the responses of five students to all questions on the examination. That approach provided us with reliable data for Mathematics A and B. We also obtained information about students' choice of other subjects, so we were able to distinguish some subgroups within the group taking the examination.

Mathematics A, mathematics in realistic contexts, was an entirely new curriculum and differed very much form what teachers and examiners were familiar with. We will, therefore, focus on it. The first examinations for pre-university education took place in 1987, so we had four years of data available for study, 1987–1990. Some problems from the examinations are shown later in this paper. The examples include p-values, e.g., the percentage of the mean score from the maximum score for the whole group. In the next section the examination results of three subgroups are compared. The development and use of a test grid for final examinations follow. We then describe some trends in mathematics education and assessment in The Netherlands. Finally, we come to some cautious conclusions.

2. RESULTS AND SUBGROUPS

Each year we obtained the results of the Mathematics A examination from a sample of more than 2,000 students. These data are shown in Table 1. Using these data from the four years, a decision was made as to what number of points would give a pass result, to fix the caesura (break). This was used to determine the percentage of students having an insufficient mark. The mark gained for this part of the examination counted 50 percent of the assessment of each student; the other 50 percent came from teacher-made tests. If these two figures differed too much, the inspector would try to find out the reasons or the cause. The percentage of students having an

insufficient final mark for Mathematics A was less than the above given figures.

Question	*p* values			
	1987	1988	1989	1990
1	96	71	63	91
2	91	89	74	83
3	60	68	78	96
4	53	79	43	60
5	55	35	44	73
6	79	36	46	71
7	66	26	31	40
8	61	93	77	74
9	60	67	56	64
10	68	81	25	91
11	49	69	85	89
12	54	80	71	56
13	38	27	51	54
14	29	60	78	31
15	60	78	20	42
16	60	76	27	66
17	39	71	61	47
18	51	44	26	31
19		36		67
20		51		
21		81		
22		51		
23		51		
Mean score	62	61	57	66
stand. dev.	17	17	16	17
reliability	85	83	.77	.83
caesura	49/50	51/52	51/52	54/55
% insufficient	25	32	37	26

Note: The maximum score one can get for this examination is 100 points. For his presence a candidate gets 10 points, the other 90 points are spread over the questions. The examiner has a strict set of correction rules, including the maximum number of points for each question.

Table 1 *Mathematics A results*

A student can still pass the final examination with an insufficient mark for mathematics, because in The Netherlands there is a system of compensation. Good marks in other subjects can compensate for an insufficient mark in a certain subject, i.e., to a certain extent. However, a mark lower than 4 on a scale of 1–10 can never be compensated for.

Those who are involved in the preparation of the examination papers have tried to estimate the mean score before the examination takes place by predicting the p-values of each question in one of five classes: Class I with $0<p<20$, Class II with $20<p<40$, etc. The prediction of the mean score in 1987 was 60, in 1988, 63, in 1989, and in 1990, 61. So the difference in results in 1989 and in 1990 were not predicted.

In 1987 we had some trouble with the statistical relevance of a certain problem and omissions in its context description, but we do not think that students were handicapped by these troubles. Nevertheless, in 1987–1990 we spent a lot of time preparing good problems with correct questions. In the meantime, we tried to understand the differences in the results of some *subgroups of Mathematics A students*. The subgroups were:

I. Students also doing Mathematics B,
II. Students with Physics, without Mathematics B, and
III. Students without Physics and without Mathematics B.

In the graph below, the mean p-values from these three subgroups are shown for the questions on the Mathematics A examination in 1990. About 33 percent of the students who chose Mathematics A are in Group I, 10 percent in Group II, and 57 percent in Group III.

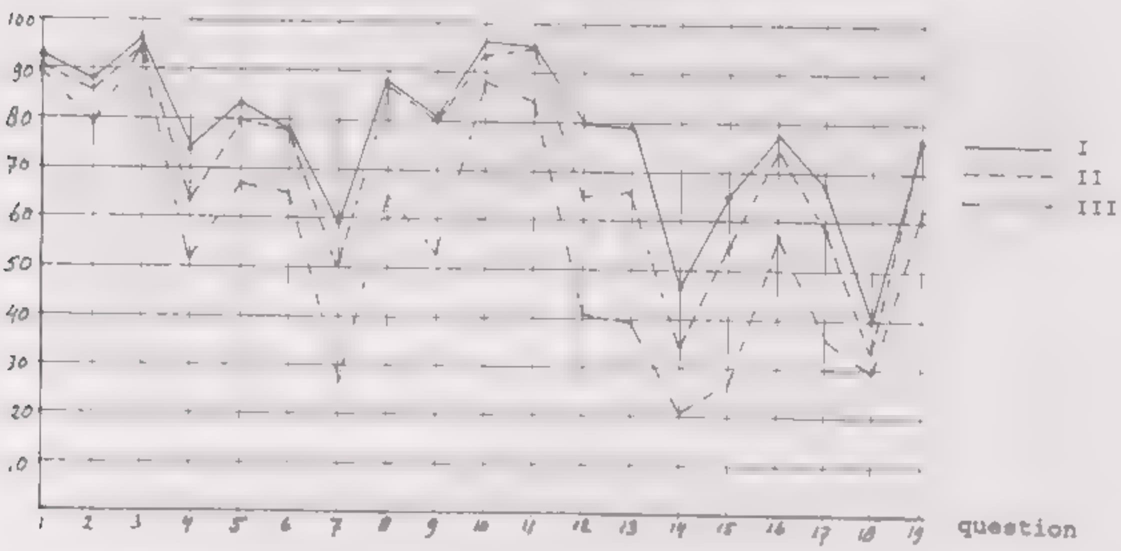


Figure 1

One must keep in mind that the Mathematics A curriculum was designed for subgroup III. Mathematics is compulsory for university studies in social sciences. But students in these subjects do not study pure mathematics as they will only use mathematics in other subjects in the future.

The mean scores for the three subgroups are:

	Group I	Group II	Group III
1987	73	64	53
1988	72	65	53
1989	69	63	50
1990	78	73	59

Table 2 *Mean scores*

The percentage of students having an insufficient mark in Group III was 41 percent in 1987, 46 percent in 1988, 54 percent in 1989 and 37 percent in 1990. This increasing percentage in the first three years was very alarming, but the result in 1990 was encouraging.

It is obvious that Group I does better than Group III because the weekly amount of hours it spends on mathematics is twice as much compared to Group III. We think that students in Group II will have an advantage in applying mathematics in physics. We were anxious to know what kind of questions caused the biggest differences. During the 1987-1990 years the questions were: linear programming and calculus in 1987; hypothesis testing, linear programming, calculus and probability distributions and population distributions in 1988; a problem containing the chainline (which has probably frightened the students of Group III), calculus and a new probability problem in 1989; periodic functions, proof in a new situation, and probability in 1990.

The disappointing results of Group III could have been caused by the fact that the curriculum was new to the teachers and they had to learn how to teach it. Indeed, during the first three years, a large group of inexperienced teachers were involved and it is tempting to blame the scores on them. But we also know that in 1989, for example, Problem 1 was not a suitable first problem, especially for Group III, because many students used calculus in computing a maximum value instead of looking at a graph. Moreover, much difficulty was evidently caused by the chainline.

3. TEST GRID

In an attempt to find better reasons for the differences between the groups and between the years, we developed a three-dimensional test grid. The next paragraph describes it. Up until now, we had used a two-dimensional test grid for the construction of examination papers. On one side we listed subject components and on the other side we listed behavioral components. Each year before starting the construction of the problems, we had agreed

upon the subject components to be tested and upon the percentage of original questions to be used. After constructing a draft examination of the paper we filled in the test grid for it. Of course this was a rather subjective activity, but if the same group of experts does the activity, it is possible to compare their results over the years. We constructed the final examination papers this way for all levels of pure mathematics. For Mathematics A, however, it was not a suitable practice. Totally different questions were fit into the same cell because we could not distinguish between questions on the same subject that asked students to do comparable activities. Therefore, we needed another component in the grid: skills.

Solving Mathematics A problems demands more than the mathematical skills required to solve equations and inequalities. One must be able to choose a suitable mathematical model fitting the context or to judge whether a given model is appropriate. After finishing the mathematical part of the problem, one must be able to transfer the mathematical results into the context in order to answer the questions raised by the problem. To judge a given model is easier than making a model. Therefore, certain skill components are needed to make distinctions among the questions.

We have grouped the components into the following main categories:

Subject components
- i. Functions, formulas, equations and inequalities
- ii. Graphs, matrices and distances
- iii. Combinatorial analysis, probability and statistics
- iv. Linear programming

Skill components
- M. To draw, make, judge, vary, and explain a model
- R. To make or finish a graphical representation of a model
- G. To read data
- W. To use mathematical skills
- T. To use a combination of, more or less, all the skills

With this new classification instrument, we were better able to analyze the final examination papers from 1987 to 1990. With the help of this test grid, we could compare the mean scores per cell for each of the subgroups. As student Group II (preceding section) covers only 10 percent of students, we have not included its results; we have compared the results of Groups I and III. We have computed the mean scores of the two groups per cell and per year and converted these figures into percentages of the maximum score.

These percentages are given in Table 3. Looking at the totals for the subject components, one can see that subject ii, matrices, had the best results, except in 1989. In that year the two matrix questions were original and were of a complex structure (skill component T). The subject, proba-

bility and statistics (iii) showed the poorest results. It is a difficult subject and one can ask a variety of questions. In this subject the questions needing mathematical skills (W) showed a good result; the students were able to learn the solving strategies. The questions dealing with modeling were all original. The three questions needing skill component G (reading data) were easy for the students. This was not surprising, but the number of questions was too low for drawing statistical conclusions.

		M		R		G		W		T		Total	
		I	III	I	III	I	III	I	III	I	III	I	III
i	'87							74	45	78	55	75	47
	'88	77	60	81	71	90	87	83	58	55	29	74	52
	'89	83	65	94	57	86	62	86	45	71	39	83	52
	'90	82	56					74	50	87	67	80	56
ii	'87	86	74					76	54			79	60
	'88	82	70					78	64			81	68
	'89									51	32	51	32
	'90	91	85							81	65	86	75
iii	'87			60	42					48	35	51	37
	'88	54	28	87	73			75	62	44	28	62	45
	'89	38	20							59	47	50	36
	'90	65	27					76	62	53	32	60	37
iv	'87	75	49							70	40	72	44
	'88	40	18							50	27	46	22
	'89							70	48	78	60	76	58
	'90									72	47	72	47
Total	'87	79	59	60	42			74	46	62	42	69	46
	'88	61	41	84	72	90	87	81	59	51	28	68	47
	'89	47	29	94	67	86	62	82	46	63	45	66	45
	'90	83	64					74	53	72	51	75	55

Table 3 *Cell mean scores as percentages of maximum score*

We see a total difference of about 20 percent between the results of Group I and III. Note that the figures at the bottom on the right hand are comparable but not the same as the mean scores displayed in the section "Results and Subgroups". In 1989 in cell (i,W) the difference was 41 percent. One can presume that the chain line problem was the cause, also, for the 27 percent difference in cell (i,R). The difference between the results of Group I and III on the subject, functions, is bigger than the differences on the subjects, matrices, and probability and statistics. This is according to the expectation, because the analysis component found in Mathematics B will help to increase the knowledge of functions in Mathematics A. Matrices and probability and statistics are not covered in Mathematics B.

The difference was exceptional in cell (iii,M) in 1990; finding a relation between mean and standard-deviation for Group III was very hard. This cell represented an original question. Looking at the behavioral components, we distinguished between reproduction questions and production questions which had some original aspects. Again we looked at the differences between Group I (students who are also studying Mathematics B) and Group III (students without Mathematics B and without Physics).

Table 4 gives the mean score as a percentage of the maximum scores.

	Reproduction questions		Production questions	
	I	III	I	III
1987	70	49	70	45
1988	70	50	63	43
1989	74	53	53	30
1990	81	67	66	32

Table 4 *Mean scores as percentages of maximum scores*

As can be expected, the scores on reproduction questions are higher than those on original production questions. But there is a trend toward increasing scores on reproduction questions over the years, while the scores on production questions are decreasing, with the exception of 1990. This 1990 change is encouraging and may be the beginning of a new trend.

4. TRENDS IN MATHEMATICS EDUCATION IN THE NETHERLANDS

In the last decade, some new trends in mathematics education have developed in The Netherlands. Up to the 1970s, innovation in mathematics

education generally meant the introduction of new subjects. So, in 1968, vector geometry and statistics were introduced into the pre-university curriculum; at the same time, solid geometry was abolished. However, during the 1970s people realized that innovation had to be more than just changes in subject matter. This was caused by developments in mathematics and also by the way people look at mathematics (Berry et al., 1984; Blum et al., 1989).

It is necessary to mention some of the important trends of assessment in mathematics here:

1. At present almost everyone comes into contact with mathematics in one way or another. Because of that, it is necessary that young people learn as much mathematics as possible.

2. For years it has been emphasized that there is a sharp difference between the two kinds of mathematics: pure and applied. Now we know this distinction is too rough. There are many nuances: from "pure", to "not-yet-applicable", to "applicable", to "applied". Moreover, whether this work is (more) pure or (more) applied often depends on the mathematician's intention.

3. One of the characteristic aspects of mathematics is deduction. In former days, we taught our pupils in school to reason in a logical way with axioms, definition of propositions, and statements about shape and number configurations. Too often we forgot that (school) mathematics was for totally different purposes. With regard to the methods used by those working in mathematics, two types can be distinguished: (a) working in an existing mathematical system, or 'closed mathematics" ("fertige Mathematik", Fischer, 1985), deduction is an important tool; (b) working open ways at new problems, or 'open mathematics" ("werdende Mathematik"); heuristics and common sense are important aspects for a sort of local deduction (van Streun, 1985).

4. When we take Points 1–3 seriously, they have consequences for the choice of subject matter as well as for the working methods and assessment in secondary mathematics education. In the 1960s and the 1970s, we thought abstract mathematical structure ought to be in the curriculum. Now it is clear that we have to pay attention to pure and applied mathematics, and that geometry and geometrical aspects have to be part of both. In the sphere of applied mathematics, we try to link up with the world of youth, and more generally with realistic situations: *mathematics in context*.

5. Volume 4 of Unesco's *Studies in Mathematics Education* (Morris, 1985) stated: "Problem-solving is being ushered in as the paramount mathematics innovation of the eighties". In the mathematics education of our country we

know are trying to realize this innovation. The "open" working method, mentioned in Point 3, can come into its own in "problem solving". Pupils can learn that mathematics is more than thinking about problems which are very simple in principle, removed from reality, and directed only toward the generalizing of an abstraction.

6. In The Netherlands much attention is paid to the position of the computer in mathematics education. Its use in situations with an abundance of quantitative data is clear. However, much research still has to be done to develop more fundamental possibilities for it in school mathematics.

5. NEW TRENDS IN ASSESSMENT

In the beginning, too little consideration was given to questions about whether the traditional ways of assessment were appropriate for the new curricula and working models. Indeed, the introduction of applied mathematics was the beginning of reflections on assessment (de Lange, 1987). In Dutch secondary education, realistic mathematics is so new that new assessment developments are mainly restricted to Mathematics A. Many mathematics teachers are convinced of the motivational aspects of drawing applications from fields outside mathematics. Assessment renewal is shaped by aspects of mathematization and model-thinking. These fit extremely well with the vision that mathematics is ultimately grounded in and connected with empiricism.

6. RISKS

It is essential that we go beyond empiricism. In empiricism school mathematics becomes a servant to the modern way of thinking about utility. Then mathematics becomes an auxiliary science only. In that way of thinking, applications would come in first place; the relevance of school mathematics would be derived from applications. Apart from the fact that this way of thinking gives rise to false — or at least one-sided — images of mathematics, it is possible that this working method would ultimately be counter productive. The excitement of mathematics is increased with the addition of attractive outside-mathematics problems. It is true that, in these kinds of problems, a more open method is being practiced; nevertheless it remains quite peripheral Not only for outside-mathematics problems, but also within mathematics itself, an open approach is possible and necessary. A teacher who does not feel the excitement of mathematics and who does not let pupils experience it, educates pupils who find mathematics tedious, for whom mathematics is only a collection of dull problems. This is one of

the main problems in mathematics nowadays, and manifests itself in a very small number of students beginning in mathematics at Dutch universities.

7. TOWARD A BALANCED WAY OF ASSESSMENT IN MATHEMATICS

Assessment in mathematics has to link up with the subject matter as well as with the working method. So the characteristics of pure and applied mathematics have come into their own in both the open and closed approaches. Taking both sets of distinctions into account we can write down the following matrix.

Approach / Mathematics	Closed	Open
Pure	1	3
Applied	2	4

Table 5 *Type-by-approach matrix*

Analyses of assessment problems in schools and examinations teach us that:

- Cell 1 is the most practiced by mathematics teachers and examiners
- Cell 2 is becoming of greater interest
- Cell 3 and Cell 4 are mostly unexploited

8. EXAMPLES

It is appropriate to illustrate the foregoing with some examples. The first example is Problem 3 of the Mathematics A examination, 1990. (The p-values appear after the questions, within bracket).

Problem 3, Mathematics A, 1990.

The BFW factory produces two kinds of cathode tubes for television sets: Square and Flat. With the present machinery BFW is able to produce a maximum of 300 Square tubes and 375 Flat tubes. BFW delivers exclusively to TV-INTERNATIONAL; according to the delivery contract TV-INTERNATIONAL will take 400 tubes per week at most. For BFW the profit on a Square tube is Dfl. 120.00 and the profit on a Flat tube is Dfl. 100.00. x is the number of Square tubes to be delivered in a week and y is the weekly number of Flat tubes.

10. Compute the value of x and y for which the profit for BFW is maximal. (91)

As TV-INTERNATIONAL has a large stock of Square tubes, negotiations took place to change the delivery contract. BFW is still allowed to deliver 400 tubes a week and the profit of the Flat tube remains Dfl. 100.00. For Square tubes the profit will become dependent on the number of the delivered tubes of this kind. If x of this kind will be delivered, the profit is Dfl. $(120\text{-}x)$ a piece.

In the figure are drawn — without the restricted conditions — some iso-profit lines for the new situation.

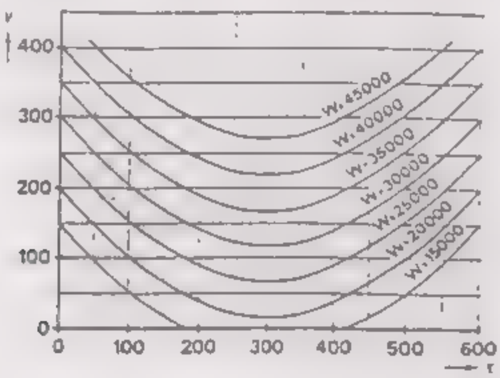


Figure 2

11. Investigate by a calculation if the points (150,150) and (250,100) are on the same iso-profit line. (89)
12. Prove that the iso-profit line $W=20000$ is part of a parabola. (56)

In a certain week, after the new delivery contract came into operation, BFW delivered the maximum number of tubes to TV-INTERNATIONAL and reached a profit of Dfl. 40,000.00. All the restricted conditions are fulfilled.

13. Compute the number of Square tubes BFW delivered that week. (54)
14. Compute the maximum profit BFW can reach in a week, considering the restricted conditions, after the new delivery conditions are introduced. (31)

Comments

In the test grid these problems were placed in Cell (iv,T), (i,T), (i,M), (iv,T) and (iv,T) respectively, while the last three questions had some original aspects. The whole problem completely fits into Cell 2 of the matrix. The problem is closed and each question only allows one answer. The word "investigate" suggest an openness which is not present at all. The problem tests knowledge and ability of the students in the subject.

The second example is from the Mathematics B examination, 1990.

Problem 3, Mathematics B, 1990.

For every $p \in \mathbf{R}$ the function $f_p : x \to x + \sqrt{1-px}$, $x \in \mathbf{R}$, is given. In a rectangular Cartesian coordinate system Oxy K_p is the graph of f_p. In the figure K_1 is sketched, A is the boundary point, B is the vertex, C is the point of intersection with the x-axis.

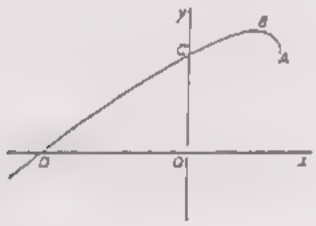


Figure 3

11. Compute the coordinates of A, B, C and D. (76)
12. Compute p if K_p and K_1 intersect at right angles. (30)

Let the differential equation

$$D: \frac{dy}{dx} = \frac{x^2 - y^2 + 1}{2x(x-y)}$$

be given. In Figure 4 a part of the field of directions of D is sketched. On the basis of the figure one can suppose there are two linear functions which satisfy D.

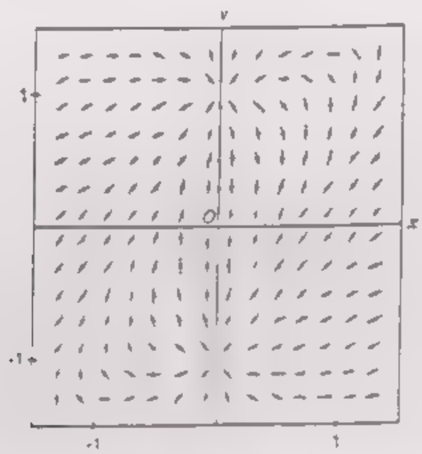


Figure 4

13. Investigate if this conjecture is correct. (45)

The given functions f_p are solutions of D.

14. Prove this. (32)

Comment

Although the questions 11 and 13 had some original aspects, the whole problem fits completely into Cell 1 of the matrix.

These comments do not suggest that the problems are wrong or improper. On the contrary, with these questions one can test whether students have or do not have the intended knowledge and abilities at their

disposal. However, good mathematics education and good assessment practices have to go further. These questions are part of a central examination. It may not be possible to go further in the situations illustrated here.

An example of a problem that shows what is meant by Cell 4 in the matrix follows. It was in the 1990 Mathematics A-lympiad (an olympiad devoted to Mathematics A in our country).

The struggle against shoplifting, Mathematics A-lympiad, 1990.
Shop-keepers at last want to make an end to shop robbery. Of course there were always customers who took little care in paying, but at present goods for thousands of guilders are stolen from shops. The thieves go from shop to shop. Many of them come from a town in the neighborhood.

There were talks with the police which resulted in a plan for a telephone warning system to be made. Fifteen shop-keepers will participate. The shop-keeper who points out a suspected person, calls two colleagues and gives a description of this person. Ringing up colleagues takes two minutes. After two minutes one of the two colleagues has been rung up, so that he in his turn can warn two colleagues, on the average after four minutes he has warned the second colleague, etc.

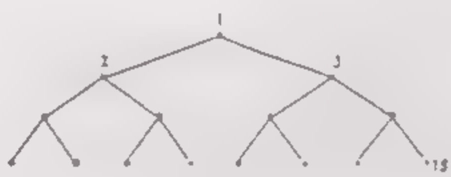


Figure 5

In Figure 5 the fifteen shops are placed in a phone-tree. If shop 1 at time $t=0$ begins to ring, after eight minutes 73% of the shops will be reached and shop 15 as the last one will be warned after twelve minutes. However, a thief will not always begin at the top of the phone-tree (shop 1) with his raid. Therefore, the phone-tree has to be adapted in such a way that every shop can start the warning system.

Task 1.
Invent various variants for this phone tree. Use as criteria for the quality of the variant the number of minutes which is necessary for everybody to be warned, and the number of minutes which is needed to warn at least half of the colleagues.

To give the shops their places in the warning system one has to take into account their relative positions and the possibility of robbing a particular shop (the theft possibility). For instance, if the personnel of a record shop identifies a suspected person, they will first ring up another record-shop and big stores and not the clothes-shop across the street. (A map of the town and a survey of the fifteen shops are provided for the students).

Task 2.
Invent a method to determine the place of the shops in the warning system. In so doing you have to take into account the walking distance between two shops and the "distance" concerning the theft possibility. Give the fifteen shops their places in the phone-tree which was the best in Task 1. Investigate if, according to the new criteria,

a totally different structure would be better. Bear in mind that each shop-keeper must be able to handle the system in an unambiguous and simple way. Write a documentation of the system, remembering that the police will be needing a total survey of the system and that each shop keeper only has to know what he has to do. Also write a "popular" story in which you explain for shop keepers how the system roughly works and why it is a good system.

In the foregoing, examples of Cells 1, 2, and 4 from the matrix were shown. There are hardly any examples, at least at secondary level, for Cell 3. It is worthwhile, however, to try and develop such questions.

9. POSSIBILITIES

In The Netherlands, school examinations consist of two parts: (1) a central part, the same for all schools of a certain type; (2) a school part, composed by every individual school. In the school part, teachers can pay attention to the aspects that are, in their view, important. Teachers can herein ask facets meant to fit the various cells (1 to 4) A more open assessment than the one which is practised traditionally is quite possible in classroom practice. We have to find other forms than those wherein every pupil works for one or more hours at the same problem. We have to, for example, consider group work: For some time (several weeks perhaps) a group of students will work in an open way at a problem. Discussions with each other and with the teacher, research in the literature, and presentations in written as well as in oral form are extremely valuable elements. Some use of this way of assessment is already underway in our country and the outcomes are highly satisfying.

10. CONCLUSION

It is useful to collect data on examinations. Interpretation of these data is often very difficult: one must collect data during a long period in order to give meaning to the results. For the construction of examinations papers it is useful to develop a test grid. Prediction of the difficulty of an examination is very difficult to do, but is always worth trying.

It is of great importance to assess mathematics in ways that include a good balance between an open and a closed approach In this way, assessment can become more adequate for the type of mathematics at issue, as well as for the method of working, and for the elements that the student most know or be able to do. It gives a better feedback to the mathematics teacher as well as to the student. Last but not least, it is highly motivating for students as well as teachers, and it contributes thereby to the solution of the biggest problem of mathematics education today.

REFERENCES

Berry, J.S. et al., 1984, *Teaching and applying mathematical modelling*, Ellis Horwood, Chichester.

Blum, W. et al.: 1989, *Applications and modelling in learning and teaching mathematics*, Ellis Horwood, Chichester.

Fischer, R. et al.: 1985, *Mensch und Mathematik*, Bibliographisches Institut, Mannheim.

de Lange, J.: 1987, *Mathematics, insight and meaning*, OW & OC, University of Utrecht.

Morris, R. (Ed.): 1985, *Studies in mathematics education, 4*, UNESCO, Paris.

van Streun, A.: 1989, *Heuristisch wiskundeonderwijs*, University of Groningen.

Wim Kleijne
Inspector of Secondary Education,
Apeldoorn

Henk N. Schuring
Institute for Educational Measurement (CITO),
Arnhem

The Netherlands

MAX STEPHENS & ROBERT MONEY

NEW DEVELOPMENTS IN SENIOR SECONDARY ASSESSMENT IN AUSTRALIA

1. INTRODUCTION

Expanding the range of performance assessed in mathematics in order to reflect more fully the objectives of the mathematics curriculum is not likely to be achieved without major reforms in the design of the mathematics curriculum, at system and school levels, and by ensuring that assessment procedures are driven by the curriculum, not the other way around.

This paper examines recent changes in curriculum and assessment at the senior secondary level in the state of Victoria, Australia, and contrasts the impact of these changes with that of previous and more limited attempts at change. Also discussed are the agents of change and the interaction between curriculum change at the lower year levels and system-wide changes in the structure of assessment and certification at senior secondary level.

2. THE AUSTRALIAN SCENE

The "New Mathematics" arrived in Australia in the early 1960s. While changes to content were at the time seen to be most important, these changes were essentially grafted on to existing arrangements for assessment at the upper secondary school.

Elements of school-based assessment, often joined to moderation procedures, were allowed, particularly for those courses where students were not expected to undertake tertiary studies in mathematics. For those who were oriented towards tertiary level studies in the mathematical sciences, the principal forms of assessment were written examinations. Since these examinations were also used to provide students with a tertiary entrance score, it was customary for examination papers to include questions ranging from relatively simple, which most students could be expected to complete, to more complex questions which would sort out the brighter and better prepared candidates. These examinations usually took the form of two- or three-hour written papers. In Australia, there has been no tradition of oral examinations in school mathematics.

M. Niss (ed.), Cases of Assessment in Mathematics Education, 155–171.

The most visible and lasting changes in secondary mathematics grew out of a multiplicity of what were, for the most part, small-scale changes in the secondary years. Innovative approaches to the teaching and learning of mathematics occurred through individualized learning, applications and problem solving. These became embodied in curriculum packages such as *RIME* (*Reality in Mathematics Education*, 1984), and the *Mathematics Curriculum Framework for years P to 10* (1988), developed in Victoria, and national projects such as *Mathematics at Work* (1988) and the *Mathematics Curriculum and Teaching Program* (1988). These and other related initiatives, though their adoption was never widespread, constituted a challenge to the more slowly changing courses and assessment practices in the upper secondary years.

Efforts to reform the senior secondary curriculum never successfully challenged the established framework of certificates, examinations, and procedures for tertiary selection. In response to growing retention rates at secondary schools during the early 1980s, there was a proliferation of alternative courses, with different content, different assessment methods, different student clienteles, and different levels of acceptance in the community, and for tertiary entry.

A Culture of Change

By the end of the 1980s, sweeping changes to curriculum and assessment in the senior secondary years were introduced in response to a changed agenda for secondary schooling, in all Australian States, where it is now assumed that the vast majority of young will complete Year 12 before moving on to further study, work, or a combination of both. As a result, there has been an extensive top-down reform of curriculum and assessment arrangements across all subjects in the senior secondary years. In mathematics, these reforms have given effect to many of the changes which had been urged for many years by the mathematics education community.

In Victoria, the *Victorian Certificate of Education* (*VCE*), introduced in all schools in 1990, is a common credential for the final two years of secondary school, based on a comprehensive curriculum for all students, comparable assessment practices for all students in Year 12, and government targets for increased retention rates to the end of secondary schooling. While mathematics is not a compulsory subject, over 90% of students in Year 11 attempt at least two units of mathematics, with a large proportion attempting four units, and a smaller proportion attempting six and eights units over the two years.

3. THE NEW STUDY DESIGN IN MATHEMATICS IN VICTORIA

The study design in mathematics provides a framework within which

teachers develop courses in *Change and Approximation* (C & A), *Space and Number* (S & N), *Reasoning and Data* (R & D) and extensions as units 1 and 2, or as units 3 and 4. Units 1 and 2 of the first three subjects are typically taken in Year 11, and units 3 and 4 and Extensions are done normally in Year 12. At each stage, a mix of core content and options is required, with approximately 50% core content in each subject. A typical 8-unit student program over the two years of the VCE could be:

Year 11	S&N units 1/2	C&A units 1/2
Year 12	R&D units 3/4	Ext. C&A units 3/4

Work Requirements and Satisfactory Completion

Work requirements and their satisfactory completion are key features of the *Mathematics Study Design*, and for all VCE courses. Work requirements are intended to be used by teachers in planning and managing each course, and to provide a clear link between what students do in each course and how their work is assessed.

Three work requirements apply to each course of study in mathematics:

- *Skills practice and standard applications*: the study of aspects of the existing body of mathematical knowledge through learning and practising mathematical algorithms, routines and techniques, and using them to find solutions to standard problems.

- *Problem solving and modeling*: the creative application of mathematical knowledge and skills to solve problems in unfamiliar situations, including real life situations.

- *Projects*: extended independent investigations involving the use of mathematics.

At least 20% of class time must be devoted to each work requirement. Each unit of the mathematics study consists of approximately 100 hours, of which 50–60 hours is expected to be offered as class time.

To quote from the Mathematics Course Development Support Material,

> It is on the basis of a student's completion of the work requirements for a unit that a judgment of satisfactory or non satisfactory completion will be made and thus whether or not the student will receive credit for that unit towards the award of the VCE. ... They allow only for a judgement as to whether the work was completed as specified in the study design. (CDSM, 1990, p. 75)

Levels of student performance, on the other hand, are determined by *Common Assessment Tasks (CATs)*.

Satisfactory completion is thus a judgement about the work a student has completed rather than about the standard achieved, although it is to be noted that the work requirements have been designed in such a way that a certain quality of work must be attained before the work requirements can be said to be completed. For example, in meeting the work required in Skills Practice and Standard Applications, students must submit "properly finished work", and may be required to resubmit work which has not been properly completed.

Work requirements are a means of helping students to meet the objectives of a unit. They are not assessment tasks, although the two are linked. Assessment tasks may be the product of all or part of the work required. In Year 11, for example, the report of an investigative project or the written report of a problem-solving task or selected assignments for Skills Practice and Standard Applications may be used as a basis for assessing levels of performance (CDSM, 1990, p. 77–78).

In summary, the three work requirements of the mathematics study are intended to give expression to a broadened range of mathematical activity by ensuring that significant time is spent on all objectives of the course. The work requirements are directly linked to criteria for satisfactory completion, and, in all units, assessment tasks assess learning that students have completed through undertaking the work required.

These links between work requirements and the range of performance assessed in the VCE illustrate the two principles discussed by Malcolm Swan (1991), the first principle being that of *curriculum balance*,

"The assessment package must consist of a set of tasks of varying length which, taken together, reflect our curriculum objectives in a balanced way",

and the second that of *curriculum validity*,

"The assessment tasks themselves must represent learning activities of high educational value so that significant time spent on them will represent a benefit rather than a loss to pupils' learning".

4. COMMON ASSESSMENT TASKS (CATS)

The new curriculum in mathematics for Years 11 and 12 requires assessment of investigative projects, problem solving and analysis, as well as continued assessment of basic concepts and standard applications for all students. In Year 11, schools are responsible for carrying out assessment tasks based on the work requirements. In Year 12, assessment is carried out through four externally designed instruments called Common Assessment Tasks (CATs).

The CATs have been developed in order to be accessible and challenging to students with a wide range of abilities; and to be appropriate to the

range of different courses which can be formed within the study design. They are also intended to constitute a reasonable workload for teachers and students, possess public credibility and equal esteem, and be informative for tertiary selection and for employers. The four CATs are common to all students taking courses at units 3 and 4. The following is a brief summary of the four CATs.

CAT 1 is an *investigative project* on a centrally set theme. This assessment task is to be carried out during a designated period, using a mix of in-class and out-of-class time.

CAT 2, the *challenging problem*, is chosen from four centrally set problems for each course. This assessment task is undertaken over a two week period, with half of the time required to be spent in the classroom.

CAT 3 is a *multiple-choice test*, of ninety minutes duration.

CAT 4 is a test, also of ninety minutes, consisting of about four *structured questions* which lead from routine to non-routine aspects of a problem.

The last two CATs are conducted at the end of the year under test conditions. The four CATs together are intended to constitute a broadened and more appropriate range of assessment in mathematics. For CATs 1 and 2 which are not carried out under test conditions, there are agreed criteria for assessment, and a verification process which involves a comparison of grades across several schools, across groups of schools, and across the State. The verification process, to be discussed later, is intended to ensure credibility and community acceptance of assessments. Brief comments follow on each of the Common Assessment Tasks.

CAT 1 Investigative Project

This tasks requires each student to carry out an independent investigation and to communicate the results of that investigation. This task is completed during a designated twelve-week period as part of the project work requirement for Unit 3.

CAT 1 is intended to enable students to demonstrate their ability to carry out an extended piece of independent work in mathematics. Students undertake a project based on a single theme set for that year by *VCAB* (the *Victorian Curriculum and Assessment Board*) in each subject. Students are given the opportunity to work on their project during class periods. The total time spent on the task is expected to be between 15 and 20 hours, although teachers and students find it difficult to keep within these

recommended limits. Each student is required to submit a report of about 1,500 words on the mathematical aspects and results of the project.

The following example (Figure 1) is taken from the Investigative Project used in Reasoning and Data in 1991 on the theme of simulation. The excerpts below contain a statement of the theme, general advice and instructions to students, and some possible starting points. Students also received additional information on the prescribed conditions for this CAT and advice on the format to be used in their project report.

During class periods, and other times if applicable, teachers discuss with students their progress on the task and are available for consultation. Students are expected to discuss with their teacher their choice of topic and how it focuses on the theme of the course; their project plan; at least the first draft of their report. This is to ensure that students are seriously engaged in the project and to provide evidence so that the teacher can attest that the work is each student's own.

Students are not permitted to work in groups except where the project requires more work (e.g., more data collection) than can reasonably be done by an individual student, or where some specific interaction between students is required.

Initial grades for CAT 1 are determined within each school according to detailed criteria which are provided for the assessment of students' written reports. These grades are then subjected to a verification procedure.

CAT 2 Challenging Problem

This task requires each student to undertake a problem-solving and/or modeling activity and prepare a report on the task. This assessment task is completed as part of the Problem Solving and Modeling work requirement for Unit 4.

This assessment task takes place early in the second semester. Students undertake a problem-solving and/or modeling task (referred to here as a "problem") selected in consultation with the teacher from a list of four externally set problems, and are required to complete a written report of their solution according to the format specified. Students are given two weeks for this task, and are expected to spend a total time of between six to eight hours on the task, including four to six hours of class time. For assessment, students must complete a separate report based on their own work. They may discuss the challenging problem with others, but any ideas obtained as a result of such discussions must be acknowledged.

CAT 2 is intended to enable students to demonstrate their ability to read and understand a problem, formulate and interpret problems mathematically; use problem-solving and/or modeling strategies; try simple cases; find patterns and formulate hypotheses; simplify complex situations; define important variables; find proofs and explanations; and interpret solutions.

THEME: Use simulation to solve a problem involving chance

Your project must use simulation to solve a problem involving chance. Your project should use mathematics that is appropriate to the focus of Reasoning and Data *or* Extensions (Reasoning and Data). You must follow the instructions given below and report on each of the specified steps in the main text of your report.

General Advice

Simulation is a useful method for investigating problems. When the problem involves chance, the simulation will involve a random experiment. The simulation process is as follows.

- i. specify the problem carefully
- ii. identify the important mathematical relationships
- iii. find, use and test a model which represents the important features of the situation

You need to check that your simulation is realistic, giving answers that agree with a real life situation. By working with a model, you should be able to investigate aspects of the real situation. It is important to evaluate the reliability of the answers obtained from your simulation.

Acknowledge which area of mathematics you use: probability, statistics, logic and [illegible]. You may choose to develop a computer program to assist you, or you may use a recognized computer package, but remember to include your own analysis of the problem.

Starting Points

You may investigate any topic related to the theme *Use simulation to solve a problem involving chance*. You must discuss your choice of topic and how it relates to the theme with your teacher and you must follow the instructions above. The examples given below are possible starting points, although it is *not* compulsory to use these ones.

Use simulation to investigate

- how many cereal packets you would have to buy to get a complete set of cards
- the chances of winning a finals series given a particular position in a preliminary competition
- stock control and inventory control (for example, the shoe shop)
- the chance of getting two identical birthday cards at a child's party
- the likelihood that two people in a class have the same birthday
- winners of horse races, or outcomes of bets for punters or bookmakers
- the variation in time to travel from A to B by public transport
- how much time you should allow to drive from A to B to arrive by 9:00 am
- how much better for clients, in a bank or Medicare office, is a single queue compared with a multiple queuing system
- the likelihood of winning simple games using various strategies
- winning prizes in a gambling game
- the chance of winning a tennis game after being two sets down.

Figure 1 *CAT 1 sample task*

The following two challenging problems (Figures 2 and 3) are taken from CAT 2 for Space and Number in 1991. Four problems were presented, and students were to choose one.

Easter Sunday

In theory, Easter Sunday occurs on the first Sunday after the Paschal full moon, which is the first full moon in Jerusalem after 21 March. In practice, the scheduled date of Easter Sunday in each year is determined by a formula specified in Christian literature.

A simpler formula was derived by the mathematician C F Gauss (1777–1855), which gives the same date as the scheduled date for every year this century except for 1954 and 1981.

The formula derived by Gauss involves the use of the symbol a mod b which means the remainder when a is divided by b.

For example, 18 mod 7 is equal to 4 since 18 divided by 7 gives 2 with a remainder of 4.

Gauss' calculation of the date of Easter Sunday is as follows.

- For the year which is x years after 1900 (for example the year 1931 has $x = 31$), the first full moon occurs c days after 22 March where
  $c = [19\ (x \bmod 19) + 24] \bmod 30.$
- The following Sunday, Easter Sunday, occurs d days after the full moon where
  $d = (2a + 4b + 6c + 3) \bmod 7$
  with $a = x \bmod 4$, $b = x \bmod 7$, and c defined as before.

You will need to make use of the following information in order to answer the questions below.

- In 1990 the full moon occurred 24.07 days after 22 March. This was a Sunday.
- The time between two full moons is 29.53059 days.

Use the Gauss formula to calculate the date of Easter Sunday for each year in the period 1990–1999.

Explain the reasoning underlying the formula for c relating it to the full moon cycle.

Now let c be any number of days, for 0 to 29 inclusive, after 22 March. Show that the first Sunday after this date occurs in a further d days, as given by the Gauss formula. In the case for which this date is already a Sunday, show that the Gauss formula gives $d = 0$.

Figure 2 *CAT 2 sample task*

Teachers are permitted to give general advice on the initial selection of a problem by students and on general problem solving strategies, but are to refrain from giving hints towards the solution of a particular problem. As for CAT 1, they monitor students' work by sighting plans and drafts during the period allowed, and keep a record of what has been seen. Students are required to retain all rough notes, and to demonstrate through

Area and perimeter

The following question was posed to a group of mathematics students

> "Are there any shapes for which the numerical value of the length of the perimeter is the same as the numerical value of the area?"

One student quickly saw that a square is a shape with this property because a square which has a side length of 4 units has a perimeter of 16 units and an area of 16 square units. The student could also easily show that there could only be one square with this property.

After looking at families of shapes like triangles, circles, rectangles and other polygons the students made the following conjecture.

> "For every family of shapes there is at least one of these shapes for which the numerical value of the area and the numerical value of the perimeter are the same."

By 'family of shapes' the students meant all shapes which are similar to a given shape. For example there is only one family of squares, but there is an infinite number of families of rectangles.

You are required to find the following.

- For which shapes does the conjecture hold?
- For each class of shapes for which the conjecture holds, give a method for finding an actual shape for which the numerical value of the area is the same as the numerical value of the perimeter

Figure 3 *CAT 2 sample task*

discussions and through work done in class time that their work on the problem and their report is their own.

Final grades for CAT 2 are determined and verified as for CAT 1.

CAT 3 Facts and Skills Tasks

This tasks is a set of 49 multiple-choice questions covering all content areas of a course. It is intended to assess students' knowledge of mathematical concepts, their skills in carrying out mathematical algorithms and their ability to apply concepts and skills in standard ways.

This Common Assessment Task is externally graded. Grades are based on the extent to which students are able to demonstrate knowledge of mathematical concepts and facts in the compulsory strand and in the optional clusters they have chosen.

CAT 4 Analysis Task

This task requires students to attempt between four and six short-answer questions involving multi-stage solutions of increasing complexity. It is

intended to assess students' abilities in interpretation and analysis of the mathematics defined by the compulsory strand of content for each course. CAT 4, the Analysis Task, is conducted under formal examinations conditions.

The following question is taken from CAT 4 for Extensions (Change and Approximation).

A batch of a particular plastic is commercially produced by mixing the ingredients in a vat. The ingredients combine slowly to form the plastic in such a way that the quantity of plastic produced, x kg, in the vat t minutes after the ingredients are mixed is given by

$$x(t) = 5\log_e(10t+1), \quad t \geq 0.$$

The cost of the ingredients for each batch is \$2,000. The cost of operating the vat for t minutes is \$$20t$. Each kilogram of the plastic can be sold for \$100.

a What quantity of plastic, to the nearest 0.01 kg, is present in the vat after ten minutes?

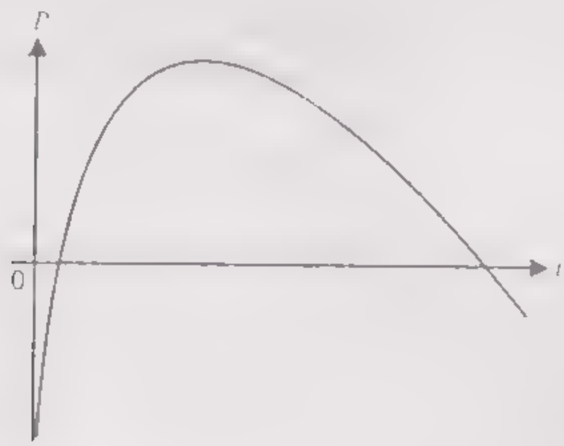
b. What is the total cost of production, to the nearest dollar, if the process runs for ten minutes?

c If the plastic produced in the first ten minutes from one batch of ingredients is sold, what profit, to the nearest dollar, is made?

d If \$$P$ is the profit from each batch of plastic, show that

$$P(t) = 500\log_e(10t+1) - 2000 - 20t.$$

e The following sketch shows the general shape and position of the graph of P against t.



i Find the time for which the process should run to maximize the profit.

ii. What is the value of its maximum profit to the nearest dollar?

f. Use approximation methods to find the shortest operating time, to the nearest minute, if a profit is to be made.

g. The cost of ingredients may vary from the $2,000 given above. What is the greatest cost of ingredients, to the nearest dollar, that still enables a profit to be made?

Figure 4 *CAT 4 sample task*

These questions require far more extended mathematical analysis than the one- or two-step questions contained in CAT 3, the Facts and Skills task. CAT 4 is designed to assess the extent to which students are able to demonstrate understanding of the required mathematical concepts and skills from the areas studied, and their use of problem-solving strategies to interpret, analyse and solve routine and non-routine aspects of problems.

Assessment Criteria and Grade Descriptors

For CAT 1 and CAT 2, teachers are required to discuss the assessment criteria with students as they engage in these assessment tasks. For the Investigative Project, the criteria are grouped under three major headings as follows:

Conducting the investigations

- identifying important information
- collecting appropriate information
- analyzing information
- interpreting and critically evaluating results
- working logically
- breadth or depth of investigation

Mathematical content

- mathematical formulation or interpretation of problem, situation or issues
- relevance of mathematics used
- level of mathematics used
- use of mathematical language, symbols and conventions
- understanding, interpretation and evaluation, of mathematics used
- accurate use of mathematics

Communication

- clarity of aims of project
- relating topic to theme
- defining mathematical symbols used
- account of investigation and conclusions
- evaluation of conclusion
- organization of material

The assessment criteria are in turn linked to *grade descriptors*. For CAT 1, these are:

A. Clearly defined the investigation and evaluated the conclusions. Demonstrated high-level skills of organization, analysis and evaluation in the conduct of the investigation. Used high level mathematics appropriate to the task with accuracy. Communicated results succinctly in the specified project report format.

B. Clearly defined the investigation. Demonstrated skills of organization, analysis and evaluation in the conduct of the investigation. Used mathematics appropriate to the task with accuracy. Communicated results clearly in the specified project report format.

C. Defined the investigation. Demonstrated some facility in the collection and analysis of appropriate information. Used mathematics appropriate to the task. Communicated the results in the specified project report format.

D. Defined the investigation. Identified and collected appropriate information. Used mathematics relevant to the task. Communicated the results in the specified format.

E. Stated a project topic relevant to the theme. Identified basic information. Used mathematics relevant to the task. Communicated the report in the specified format.

Each of the assessment criteria is to be rated on a four-point scale: High, Medium, Low, and Not Shown. The criteria are directly related to the grade descriptors. These give a ten-point scale ranging from A to E, with two levels within each, that is A+ as well as A. Two further "grades" are available Ungraded (UG) and Not Assessed (NA).

Samples of reports completed by students have been used as a major focus of the VCE teacher development program to help teachers become more confident in judging how to apply these criteria. At the heart of the verification process are teachers' explanations of how they have applied the criteria to their students' reports for CATs 1 and 2.

For CAT 2, the Challenging Problem, assessment criteria are also grouped under three major headings. These are:

Defining the problem
- clear definition of what is required
- definition of important variables, assumptions and constraints
- identification of nature of solution sought

Solution and justification

- production of a solution which addresses the problem
- degree of mathematical formulation of problem
- appropriate use of mathematical language, symbols and conventions
- accuracy of mathematics
- interpretation of mathematical results
- depth of analysis of problem
- quality of justification of solution

The solution process

- usefulness of questions asked
- relevance of mathematics used
- generation and analysis of appropriate information
- recognition of the relevance of findings
- refinement of definition of problem

The assessment criteria can be used as signposts for teaching. If, for example, students are to appreciate the importance of defining a problem in terms of important variables, assumptions and constraints', these elements of problem-solving should be discussed by the teacher and supported by a careful choice of practice examples or samples of students' work. Teachers plan activities which sharpen students' understanding of the assessment criteria.

Impact of Work Requirement and Associated CATs

The introduction of work requirements and associated Common Assessment Tasks has had a significant impact on the teaching of mathematics in Years 11 and 12, and also in preceding years of secondary school.

The public nature of the assessment criteria for all four CATs directly links the work of teachers and the work of students to the assessment process. Teachers must ensure that the criteria for satisfactory completion of work requirements are met, and that students understand and are capable of meeting the expanded criteria for assessment.

The organization of mathematics classes has changed to take account of the requirement for consultation and joint planning between teachers and students, in for example, the initial discussion of the project or problem in CAT 1 or 2, monitoring its development, and examining first drafts of reports.

In their turn, students need to develop skills in the writing up of reports on problem solving and modelling tasks, as well as writing extended reports on investigative projects. Students also require access to a wider range of teaching resources, including access to data sources and information from other disciplines.

Impact of the Verification Process

Teachers' decisions about the award of grades have to be defensible to their students, and credible across schools. The verification process requires teachers within a school to confer with one another about the application of the assessment criteria, and subsequently to share their assessment with teachers from other schools. This is to ensure that the assessment criteria are consistently applied by each school, and also to ensure the credibility of assessments themselves and their comparability across schools.

These interactions among teachers which are the heart of the verification process enable teachers to see a direct link between their own practice and learning outcomes for students. In addition, there is substantial and independent re-grading of students' work sampled from each school by chairpersons of verification panels whose special role is to maintain standards and to see that they are applied consistently. This re-grading is of particular value to teachers in adjusting their initial assessment.

Impact at Year 11 and Below

Schools are responsible for all aspects of assessment of levels of performance in courses taken at units 1 and 2. The *Course Development Support Material* (CDSM) provides advice on the design of assessment tasks which might be used to provide assessments of students' levels of performance, and how these might be reported by means of grades or descriptive comments. In practice, schools have tended to use a range of similar assessment tasks and criteria as used for units 3 and 4.

The requirement of an expanded range of assessment tasks in the two senior years had an immediate impact on the teaching of mathematics in the junior secondary years. Extended problem-solving tasks have been introduced widely, with some teachers also introducing small-scale investigations. Many teachers tend not to make a hard-and-fast distinction between the two, preferring to concentrate on extending students' ability to work through non-routine problems and to justify their solutions.

A stronger emphasis is placed on having students become familiar with a range of problem-solving techniques, such as creating a table, looking for exceptions, using diagrams; and to develop their communication skills by writing short reports of their investigations. In assessing these reports, teachers, not surprisingly, have tended to modify and simplify the assessment check-lists which are used in the senior years. The introduction of investigative work has also given a renewed importance to areas of mathematics, such as probability, statistics and elementary mathematical modeling, in the junior secondary years.

These changes in the junior secondary years are not simply the result of changes at the top. Greater emphasis on problem-solving, for example, has been consistently advocated by all States and Territories in their advice to

schools, and has been strongly endorsed by the national statement on Mathematics for Australian Schools (1990). However, for many teachers in the State of Victoria, the VCE has required a change of practice which these other policy pronouncements have merely encouraged.

Impact on Tertiary Education

The VCE's intended to provide scope for curriculum diversity, while retaining the power of externally designed assessments set in such a way as to be accessible to the whole range of students. Tertiary institutions have generally welcomed the broadened range of assessment activities in the VCE. However, their concerns have been most evident in discussion about the reporting scale proposed for the CATs. During the CATs trails, VCAB used a five-point scale with grades: A, B, C, D, E.

At issue has been whether this scale is fine enough for selection into tertiary courses, especially those where large numbers of students apply, and where selections need to be based on reasonable distinctions between students. While there may have been administrative advantages in a 100-point scale, which was used previously to score students' performance on a written examination, there needs to be a balance between

> a scale which is too coarse to express real differences and one which is so fine that it invites use of differences for which there is no real basis" (McGaw et al. 1990, p. 30)

Although a 5-point scale may have been adequate, aggregates based on an expanded 10-point scale are not likely to produce too many tied scores at critical points for determining tertiary entry.

Beyond Curriculum Design

The VCE Mathematics study design embodies a three-way linkage between the objectives of the course, work requirements and the range of performance assessed. Work requirements ensure that the course objectives are translated into time spent in teaching and learning. In turn, assessment tasks are closely matched to work requirements.

Widespread changes in assessment practice, however, are not the result of sound curriculum design alone. They require the active backing of government, school system, teachers and teacher unions, universities, textbook publishers, parents and the wider community, in a climate conducive to change. These ingredients do not come together easily.

NOTES

While this paper refers to development in one Australian State, related innovations in the assessment of mathematics in the senior secondary years are taking place in several other Australian States and Territories. The following references are illustrative of some of the changes.

In 1992, South Australia and Tasmania each introduced a senior secondary certificate similar to the VCE across all subjects. Each will assesses a broad range of performance in mathematics. See South Australian Certificate of Education, 1991 and School Board of Tasmania, 1990.

In Western Australia, one component of assessment in mathematics comprises externally structured tasks conducted by each school. These include conventional test items and investigations. See Secondary Education Authority of Western Australia, 1990.

In Queensland, assessment of senior secondary mathematics is totally school-based within guidelines provided by the Board of Senior secondary School Studies. Assessment in mathematics includes problem solving and modeling. See, for example, Curriculum Services Branch, 1990.

REFERENCES

Australian Academy of Science: 1988, *Mathematics at Work: Modelling your world*, Australian Academy of Science, Canberra.

Australian Education Council: 1990, *Mathematics for Australian Schools*, Document endorsed at December 1990 meeting, Australian Education Council, Melbourne.

Curriculum Branch, Ministry of Education: 1988, *Mathematics Curriculum Framework*, Curriculum Branch, Ministry of Education, Victoria, Melbourne.

Curriculum Development Corporation: 1988, *Mathematics Curriculum and Teaching Program (MCTP) Professional Development Package*, Curriculum Development Corporation, Canberra.

Curriculum Services Branch: 1990, *Senior Syllabus in Mathematics and Society*, Curriculum Services Branch, Department of Education, Queensland.

McGaw, B., Eyers, V., Montgomery, J., Nicholls, B. & Poole, M.: 1990, *Assessment in the Victorian Certificate of Education*, Victorian Curriculum and Assessment Board, Melbourne.

Ministry of Education: 1984, *RIME (Reality in Mathematics Education Teacher Development Project)*, Ministry of Education, Melbourne.

School Board of Tasmania: 1990, *Mathematics Applied Level III, Draft syllabus*, School Board of Tasmania, Hobart.

Secondary Education Authority of Western Australia: 1990, *Assessment support material in mathematics*, Secondary Education Authority of Western Australia, Perth.

South Australian Certificate of Education: 1991, *Domains and components: information developing Stage 1 school programs in mathematics*, Draft version, South Australian Certificate of Education, Project Mathematics Team, Adelaide.

Swan, M.: 1991, 'Improving the design and balance of mathematical assessment', paper presented to the ICMI Study on Assessment in Mathematics Education and Its Effects, April 1991, Calonge, Spain.

Victorian Curriculum and Assessment Board: 1990, *Mathematics Course Development Support Material*, Victorian Curriculum and Assessment Board, Melbourne.

Victorian Curriculum and Assessment Board: 1990, *Mathematics Study Design*, Victorian Curriculum and Assessment Board, Melbourne.

Max Stephens & Robert Money
Victorian Curriculum and Assessment Board,
Carlton, Victoria,
Australia

LEONOR CUNHA LEAL & PAULO ABRANTES

ASSESSMENT IN AN INNOVATIVE CURRICULUM PROJECT FOR MATHEMATICS IN GRADES 7–9 IN PORTUGAL

1. INTRODUCTION

The innovative MAT 789 Project is developing a new (non-official) mathematics curriculum for grades 7–9 (pupils aged 12–15). From 1988 to 1992, four experimental classes in the area of Lisbon, Portugal, are involved. The Project Team has five members, three teachers and two researchers, and includes the authors of the present paper. The Project is independent, but is authorized by the Ministry of Education and supported by the University of Lisbon and other institutions.

Mathematics is viewed as a science that is evolving, a human achievement that is found in all areas of human activity, and can be learned by everybody. According to the perspective of the Project, mathematics education should develop in students a positive attitude of self-confidence, as well as an understanding of the role and importance of this science in his or her life, and in society. The learning process should be oriented essentially toward construction (not "absorption"), in such a way that transmission and repetition mechanisms play only a secondary role. This construction should appear naturally in appropriate contexts, the concepts being built by students as the proposed activities develop. In this sense, a given problem situation is not presented merely to motivate or to introduce a new concept, or to apply one; it provides a context for students' work.

2. THE PROBLEM OF ASSESSMENT

The new goals, methods, and learning activities prepared for the experimental curriculum pose the problem of how to assess and understand students' achievement.

Assessment and Change

The problem of assessment is not a new one but, lately, we are beginning to observe a turning point in the way we face it both in the procedures we

M. Niss (ed.), Cases of Assessment in Mathematics Education, 173–182.

use and our understanding of them. The growing importance given to assessment and the dissatisfaction with current practices are obvious in many recent documents in the area of mathematics education. For example, in the *Mathematics Counts* report (Cockcroft, 1982) we read that

> "assessment needs to be accompanied by appropriate recording of progress, whatever form of recording is used, some effort should be made to record qualities which can only be assessed by the professional judgement of the teacher, such as pupil's persistence in working at a problem, his ability to use his knowledge and his ability to discuss mathematics orally. Testing, whether written, oral, or practical, should never be an end in itself but should be a means of providing information which can form the basis of future action (p 122).

In the NCTM *Standards* (1989), a whole chapter is dedicated to assessment. There is a concern about assessment for the teaching process.

> "Assessment has been based on the assumption that students are collectors of knowledge and that the assessment process should examine, primarily in a static way, the collections of knowledge and understanding they possess", and yet "learning is not a matter of collecting but of constructing" (NCTM, 1989, p. 141).

> "The assessment of students' mathematical power goes beyond measuring how much information they possess to include the extent of their ability and willingness to use, apply, and communicate that information" (NCTM, 1989, p. 205).

A similar concern is expressed by de Lange (1987), who says that a detailed description of objectives and methodologies is necessary in order to develop tasks and adequate tests. In this sense, Hein (1980, p. 64) states that

> "the assessment process must take into account the educational environment, that is, it's necessary to adapt the assessment to the program rather than the opposite

Methods and instruments of assessment should, thus, be consistent with the teaching methods used.

What then, is the meaning given to assessment? As we read in the NCTM Standards (1987, p. 138–139), assessment should be "a continuous, dynamic, and often informal process". It must measure the efficiency of teaching, diagnose the difficulties of the students, provide the teacher with valuable information, give clues to the student about the quality of his or her work, give him or her fundamental feedback on the work, in all, play an important role in an effective teaching process In addition, the aim which has basically been to test and grade, should be extended, "its basic purpose ... to determine what and how students think about mathematics".

So, an instrument that only focuses on the right answer is not adequate; it is important to understand the student's comprehension of the mathematical ideas and processes, "an assessment which favors processes rather than products "(de Lange, 1987, p.163).

The need to change is strongly emphasized by Romberg (1991):

"Current tests reflect the ideas and technology of a different era and world view. They cannot assess how students think or reflect on tasks, nor can they measure interrelationships of ideas ... Only when new instruments are developed will we no longer be bound by old assessment procedures rooted in the traditions of the Industrial Age" (p. 23)

Principles of Assessment Adopted by the Project

The assessment procedures that have been adopted by the MAT789 Project follow some of these principles. They consider the previous concerns and ideas and the relevant experiences developed in the last years by the Dutch *Hewet Project* (de Lange, 1987).

The first principle refers to the understanding of assessment as an intrinsic part of the learning process. Assessment should not happen in special moments or at the end of each term, but along with the learning process, creating situations which help the learning process.

Secondly, assessment methods and instruments should be consistent with the principles used in instruction. Therefore, considering that the defined goals not only contemplate the cognitive aspects, but also include affective or social attitudes, assessment must also consider these areas. Other aspects still deserve consideration: The existence of a variety of situations in the learning process implies the use of different forms of assessment and therefore written and oral tasks for both individual and small group work; on the other hand, the focus on processes rather than on factual knowledge must be retained in the assessing process.

The third principle relates to a preference for positive assessment, that is, assessment that values what the student knows rather than what he or she does not know. The created situations must give students an opportunity to develop their potential without requiring the same level of performance of them.

Another assumption is that the form of selected assessment must not depend on its possibilities for quantitative scoring. For our purposes, a qualitative score is as valid as a quantitative one, and the choice is based on the one that best matches the instrument used. Qualitative should not be thought of as arbitrary, since qualitative does not mean the absence of criteria.

Finally, assessment must always take place in a clear and comfortable environment, where criticism and suggestions for the future are natural. The creation of anguish and stress should be avoided at all cost.

3. FORMS OF ASSESSMENT USED

Two-Stage Tests

Two-stage tests, inspired by the ideas of van der Blij and used by the Hewet Project (de Lange, 1987), include easy questions among others. We have been working on this kind of test with our 12–15 year old pupils. In a first stage, the test is given in the classroom with a time limit. Each pupil has two hours to work on the test and is allowed to consult his or her notebook or booklets. After this phase, the teacher takes the tests home and makes a first evaluation, signalling the more serious errors and asking questions or providing comments that might act as a clue or a challenge to the pupil's work. The tests are then given back to the pupils, thus, initiating the second stage. Now the pupil reworks the test at home, for a pre-arranged period of time. When this is over, the test is handed back to the teacher, who makes a new judgement. In both stages, the criteria applied to each of the questions were considered in addition to a global consideration of the test. The two-stage process ends with final information, based on the two scores and the student's evolution, expressed in qualitative terms. Generally, teachers of experimental classes have been using four levels (three positive and one negative) for the final purpose, as they do for most assessment tasks.

Example 1
One of the questions included in a two-stage test for 7th graders (12/13 years old) was written as follows:

> You may know three different methods to determine the greatest common denominator of two numbers. In earlier grades you have learned how to do it from corresponding sets of divisors. This year you did it in the classroom using a process based on the prime factors of the numbers. In your booklet about natural numbers another method, the Euclid's algorithm, is described.
>
> Try to answer the following question. In which situations does is seem to you more convenient to use one or other of these methods? Do the experiments you find necessary to form a personal opinion and present them together with your answer.

This open-ended question is intended primarily to stimulate our pupils to reflect on the methods they had used or they could now study about the topic. A best answer did not exist and we could hardly imagine what our pupils would write as an answer for this unexpected (probably) question. It required that students: understand the problem; follow a written description of a method not previously practised in the classroom (Euclid's algorithm); choose relevant examples; and write personal ideas about the issue.

As would be expected, some interesting answers were given at the first stage because most pupils took it as a question to be developed for the second stage. The diversity was enormous, going from simple randomly chosen examples for each method without further comments to some very elaborate answers (for this age level), for example:

> "The algorithm of Euclid will be adequate when numbers are like 'a' and 'something' times 'a' because one single division is enough, this is not case of prime numbers or others ...".

This example illustrates some of our major concerns when planning a written test: (a) to provide new opportunities to learn; (b) to consider the goals of the curriculum; (c) to accept different answers, based on varying perspectives; (d) to encourage each pupil to show what he/she can do.

Example 2
A two-stage test on functions given to 8th graders (13/14 years old) included the following question:

> Watch the following graphs carefully:

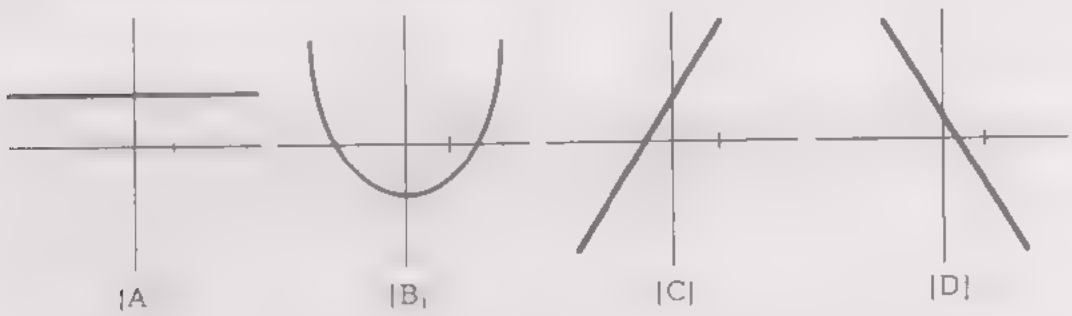


Figure 1

> Which of the graphs may represent the function of Q to Q defined by the expression
>
> $$f(x) = 3/2x+1?$$
>
> Why?

The answers can be gathered into two groups. The first include those pupils who gave numerical values and checked the correspondence between the graphs and the pairs of obtained values. Here is a pupil's answer as an example

> "Graph C is the one which may represent the function Q in Q defined by the expression $f(x)=3/2x+1$. In order to explain why I chose graph C I'll make the following scheme to demonstrate my way of thinking

x	0	1	2
$f(x)$	1	2.5	4

and these results fit the Graph C."

Other pupils gave more elaborate justifications without using numerical values, based on their previous study of functions. Nevertheless, even in this case, one can find different levels:

"The Graph B is out, because it is a parabola and to obtain one, we need and expression raised to the square, D is out, too, because we need to multiply the expression by a negative number. Graph A is also out because it is of a constant function. So, it has to be C"

The graph that may represent the algebraic expression $f(x)=3/2x+1$ is C. Because Graph A corresponds to a function where images are all alike and objects are not important, Graph B is out because it is a parabola and therefore it must correspond to a function where x has a pair exponent minus another number, in this case 1.25 or so, Graph D can't be because this straight line corresponds to a function more or less like $f(x) = x+5$ or 1. So, Graph C is the only one left and because the straight line crosses the point (0,1), that is, it was added 1 [$f(x)=3/2x+1$] and the other computations also belong to this graph but one can't see with exactness because the graph is no made on millimetrical paper

The three answers shown above were given by students in the first stage of the test. The examples illustrate a situation where pupils could choose one of various possible approaches. It is obvious to us that these answers are all correct, yet each shows different levels of learning.

Example 3

One of the parts of a two-stage test for 9th graders (14/15 years old) was to be answered by the pupils using a computer. The task required the construction of a given figure which included a square and circles of different sizes, using a computer program ("Logo.Geometria"). This program, built on the *LOGO* language, is a tool to explore problems of Euclidean geometry; the pupils had been working on it once a week for about two months.

Each pupil had 40 minutes (one third of the total time for the test) to develop this problem and had to move to another classroom to do so. This move could have been disturbing since it was a situation new not only for the pupils but also for the teachers. Reactions, however, were positive; most pupils produced meaningful answers and accepted easily the fact that the product for this part of the test was to create and save a file on a floppy disk.

Once again, some of the principles stated earlier are present, namely the congruence between learning activities and assessment tasks. Moreover, new opportunities to learn were given to the pupils. This was obvious in the

case of two girls who were obliged, for the first time, to deal with a problem situation involving the computer. On previous occasions, they tended to play a passive role, leaving to their colleagues the responsibility of the work. This situation did not have negative effects for the two mentioned pupils since a second stage of the test followed, allowing the pupils to come back to the computer and take enough time to rework and reflect on the problem.

Essay-Type Questions and Short Reports

Very often, pupils should work on and present written comments on given situations. Three examples of this kind of task follow:

- A personal comment on a newspaper article about the system of car license plates (considering the size of our population, other countries' systems and their development, as well as other aspects you may find relevant, is our system a good one for the future and what are some practical suggestions?) — Grade 7.
- A group report about the tourism conditions of a given region, considering data about temperature and precipitation for an extended period of time ("imagine you work for a travel agency and you have to write recommendations for vacations during the various months ...") — Grades 7 and 8.
- An individual report on a computer program explored by the pupils in the classroom, where they describe the general functioning of the program as well as their strategies to solve the proposed problem (to estimate the time spent by a car to go a given distance using different rates of speed) — Grades 8 and 9.

These kinds of tasks can be done individually or in groups, in a short time (for example, a part of a lesson) or during a longer period of time (for example, one week), and related directly to or independent of previous work in the classroom.

Assessment of these tasks considers both a global evaluation of the work and a set of criteria focusing on aspects such as the structure of the work, the extent to which the subject has been explored deeply and developed, the correctness of the content, the quality of the communication, and the originality of the work (if it exists). A qualitative score is given to each task and the teacher comments on it to the individual pupil or to the group, pointing out the strongest aspects of the work and also those that will require more attention in the future.

Although they have different characteristics, these tasks fit the principles of assessment stated earlier — namely, they provide new learning situations and, by their own nature, they encourage each pupil to use and develop his or her own skills and preferences. Another relevant point is the amount of

responsibility and autonomy the tasks require, a generally accepted requirement for adult activities (for example, in teacher training activities) but one seldom practised with children: In order to write a report it is necessary to organize the ideas, to start form a draft, to ask for help or suggestions, to improve the work ... One of our pupils expressed it in this way: "We have an idea and we think it is only yo write ... but we begin and then we see that there are other things to say ...".

Project Work

The previous remarks about essay-type questions and short reports hold for project-oriented activities. Additional data related to pupils' commitment and progress are obtained because projects constitute larger and longer activities, and pupils' work must be organized both inside and outside the classroom to complete them. Although different in content, projects (developed two/three times throughout the year during three/four weeks) include various stages, some of them corresponding to work in the classroom: definition of the problem and discussion on global strategies; small group work in some moments and; discussion on preliminary results. This allows teachers to see what is happening at different moments and to have information that goes beyond the contact with intermediate and final forms of a written report.

The final products generally include written (either individual or group) reports. In some cases, however, pupils build materials or models or organize an exhibition in the school. For example, the last project for Grade 7 in 1989/90 consisted of the elaboration of a detailed plan for "the ideal classroom". One week before the conclusion of the work — which included a written report and a scale model — a lesson was dedicated to oral presentations on the work in its pre-final form. This corresponds to an aspect of pupils' work seldom explored. Each group had ten minutes to make a "short presentation" and five minutes to listen and/or to answer questions or respond to suggestions from other pupils and the teacher. This constituted an opportunity: to organize and make a representation where talking about mathematical processes was necessary; to take advantage of others' suggestions to improve their work.

Throughout each project, information about each pupil and/or group — of an essentially absolute (rather than relative) and qualitative (not quantitative!) nature — was collected by the teacher and communicated to the pupils. In this sense, assessment of project work was not identified with assessment of a specific outcome. However, this did not exclude the assessment of particular products involved in a project. In the case of the previous example, the teacher assessed the report and the scale model presented by each group in a manner similar to that described above for short reports and essay-type questions.

Another example is the project developed in the same year by the 8th Grade classes. The study focused on the way the school's **cafetaria** functioned and was based on the opinions of the students of the school. The pupils developed a questionnaire, selected the sample, collected the answers, and worked on the data with the support of a computer. Finally, they discussed the main results and organized posters for an exhibition. There was no assessment of specific products, yet the work made it possible for the teacher to collect information, beyond other aspects, about pupils' attitudes, especially their sense of responsibility and personal commitment.

4. PROMISING AND CRITICAL ASPECTS

There is no doubt that the new system of assessment pleased our pupils, although in the beginning they were surprised and lacked confidence (for example, in the first two-stage test, many pupils viewed it as a kind of trick to oblige them to correct the mistakes). However, individual interviews at the end of the year showed a consensus about assessment procedures. Some pupils expressed enthusiasm about the tests. "This kind of test is like having a lesson again!", or "I adored our test of geometry, it wasn't the easiest but it was the one I preferred". Positive reactions and personal commitment were even more evident during other activities, especially in project outcomes.

Maybe more importantly, pupils showed an increasing capacity to evaluate their own work, making balanced observations about their personal involvement in different tasks. This seems to be a promising result since we expected some confusion from the loss of quantitative information from the traditional written tests.

One relevant result of our work was the positive change in the classroom atmosphere. Anxiety is very commonly associated with school mathematics, and assessment methods play a decisive role in creating anxiety. The positive nature of our assessment may lead to less anxiety.

Taking into account our lack of experience with these kinds of assessment methods, this project was also a positive experience for the teachers. It required a lot of reflection on the goals of the curriculum and the intentions and nature of each activity. In the beginning, the loss of objectivity was compensated for by the collective work of the members of the Project Team; both the researchers and the teachers expressed their opinions about each pupil, independently. This procedure allowed us to make corrections but also to become more and more confident in our ability to use the methods.

Our work is far from conclusive. Considering our rigid and centralized school system, it represents a necessary and promising innovation. At the present stage, it seems to be very important to identify the major difficulties and weaknesses of the work. We need to improve (in the sense of making

them more operational), the instruments related to the assessment of oral tasks. Although oral instruments were not absent from our work, written instruments were dominant. We need to organize a more systematic way of observing pupils' work, both during individual and small group activities. These two aspects deal with the different goals of the curriculum, but they have to do, in particular, with the assessment of attitudes which seems to be especially difficult.

A final remark relates to teachers. In our experience, while each class has a single teacher, there is a team working on the major aspects of the curriculum. We do not know what would happen in the usual situation when teachers work alone with their classes.

NOTE

The MAT789 Project Team includes the authors of the present paper and Eduardo Veloso, Margarida Oliveira, and Paula Teixeira. The Project is supported by the Department of Education of the Faculty of Sciences, University of Lisbon, and it is funded by the Gulbenkian Foundation.

REFERENCES

Cockcroft, W.: 1982, *Mathematics Counts: Report of the Committee of Inquiry into the Teaching of Mathematics in Schools*, Her Majesty's Stationery Office, London.

Hein, G.: 1980, 'L'évaluation selon la perspective de l'école ouverte', in Paquette, C., Hein, G. & Patton, M., *Évaluation et Pédagogie Ouverte*, N.H.P., Ottawa.

de Lange, J.: 1987, *Mathematics Insight and Meaning*, OW & OC, Utrecht.

National Council of Teachers of Mathematics: 1989, *Curriculum and Evaluation Standards for School Mathematics*, Reston, VA.

Romberg, T.A.: 1988, 'Evaluation: A Coat of Many Colors', in Robitaille, D. (Ed.), *Evaluation and Assesment in Mathematics Education*, Science and Technology Education Series No. 32 UNESCO, Paris.

Leonor Cunha Leal & Paul Abrantes
Department of Education,
Faculty of Sciences,
University of Lisbon,
Portugal

WEI CHAO-QUN & ZHANG HUI

EDUCATIONAL ASSESSMENT IN MATHEMATICS TEACHING: APPLIED RESEARCH IN CHINA

1. INTRODUCTION AND BACKGROUND

Fuxin, one of the cities in the west of the Liaoning Province, China, has a population of 700,000. It is a backward area in terms of economy and culture. For several reasons, education in Fuxin developed more slowly than in other cities, and the quality of the students and teachers was also less developed. According to a 1985 calculation, only 6.29 percent of all its junior school teachers had graduated from 4-year colleges, 33.72 percent from teachers normal schools, 46.25 percent from the former high schools, and 13.73 percent from the middle schools only. It is quite clear that most of the teachers do not have adequate amounts of formal schooling. On the basis of the standard of an Intelligence Quotient (IQ), the students here are below average. In 1986 in our research group, we examined the mathematics marks of the students in Grade One, and to our great surprise, only 3 percent got 90–100 (marks), 14 percent got 70–89, 34 percent got 60–69, and others who got below 60 marks made up 49 percent. (It was a hundred-mark system.) Most of the students' marks were below the country's average. For these reasons, our research group decided to start with the function of educational assessment in relation to middle school students (aged 13–15), to apply theory to practice, to improve the teaching process, and to do research on applied educational assessment.

2. THE CONCEPTION OF THE THEORY

The Purpose of the Project

What is applied assessment? It relates the level of content taught to standards, shows the important points of teaching, and judges the value of the results of the teaching process according to general aims which meet the needs of the society and the demands of the subjects so as to give teaching balance and proportion. This kind of role in activities determines its position and function in the teaching process. Educational assessment should be applied to teaching, as the pivot point, in every important part

M. Niss (ed.), Cases of Assessment in Mathematics Education, 183–191.

of teaching. For a long time, such sound uses of assessment were distorted. Tests took the place of the assessment. Teachers thought only about the number of students who would go to college, and students thought only about their marks. Marks became the only object of the assessment. This led the students to pay little attention to studying and the quality of teachers went down. We must correct this attitude before we carry out educational assessment, and return the assessment of teaching to its true role.

The Position of Assessment in the Whole Teaching Process

As a science, teaching must conform to intrinsic laws. The teachers or pupils must know why they teach and why they study (purposes of teaching and learning); what to teach and what to learn (contents of teaching and learning); how to teach and how to learn (methods and ways of teaching and learning); and outcomes of teaching and outcomes of learning (results of teaching and learning). These form the total process of teaching. The purpose of assessment is to help the educational process function fully. Assessment should be used in each of the steps such as the aims or contents of teaching, the measurement of perceptual and rational processes sought by the activities and methods of teaching, and the evaluation of the results of teaching or after-class coaching.

Assessment should, and can, strengthen the teaching function and help it to reach certain aims. Feedback and regulation need specific methods of assessment. A contemporary view of assessment is made up of aim, measurement, and evaluation. Applied assessment incorporates the contemporary view of assessment into the teaching process. The structure is depicted in Figure 1.

Such a process of teaching makes teaching and assessment act together; it forms a new system or teaching structure with the teaching aims being central and blending organically with assessment to deal with aims, activities, regulation, and development.

Basic Variables of Assessment

To strengthen the reliability of assessment and its applied value, we look on it as a supportive function that should affect every part of teaching, and we think of researching this supportive function as the most important condition, if we are to improve or regulate teaching.

The most basic requirement of our country is training enough qualified people. Everybody knows that our country is a developing country, and a qualified person is a treasure of treasures. These persons should have high moral character, be disciplined, love the motherland and the socialist cause, be devoted to hard work for the prosperity of the country and its people, seek new knowledge continuously, seek the truth from facts, think

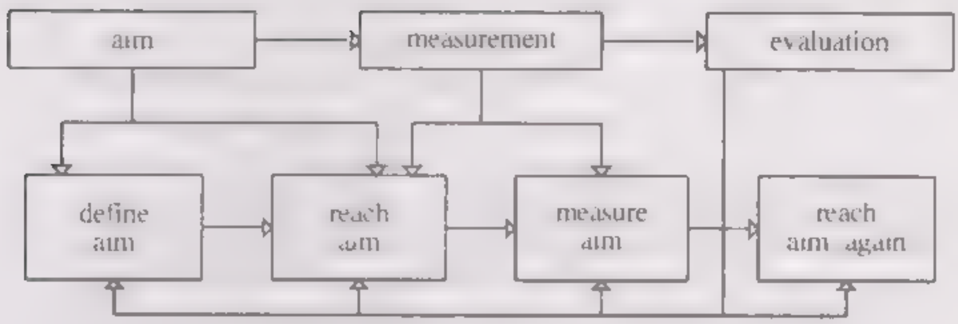


Figure 1 *The structure of applied assessment*

independently, and create bravely. Such requirements lay hard work on us. On one hand, when we give knowledge to students, we simultaneously should foster their ability and their morality. On the other hand, our teaching should foster the development of each individual. Aims are not only to obtain knowledge and techniques, but also to integrate developing knowledge, technical ability, and a scientific method of thinking. The quality and the capacity of the teaching aim and the level of students' cognitive development are a unity. Teaching programs must take the reality of students into account. There must be a definite and concrete explanation to describe an aim (including knowledge, ability, emotion, and thought) and how to reach the standards set by that aim. *Aim*, then, is one of the basic variables in assessment.

Feedback and *regulation* are other basic variables in the assessment of teaching. By using assessment to judge to what extent the teaching aims have been reached, teaching and learning are improved. Teaching and learning can be achieved in this way. A quiz before class is setting the stage to diagnose and remedy the students' difficulties before they start new lessons. Timely feedback is obtained by asking questions, discussing, doing exercises, and having exams, etc. These are important steps for finding out the discrepancy between student learning and aims, and changing the discrepancy immediately. These steps can be used to get quick feedback from students and at the same time the steps can help us to do after-class coaching. This feedback system goes on throughout the teaching process, and reaches all the students taking part in the teaching activities.

Successful regulation has many aspects. A clear teaching aim and the bilateral activities between the teacher and the students are the most basic ones. This requires that the teaching aim be apparent to the students, in fact established between the teachers and the students. The regulation and control of teaching must be timely: the teaching plan should be devised to correspond to the students' situations, the rhythm of teaching, the method of teaching, and the assignment of homework. We should also strive to promote the development of the cognitive structure of the students; to help students deal with new and old knowledge, reasoning and results; and to distinguish between matters of primary and secondary importance.

The result of teaching is another basic variable It is the conclusive factor in the system of feedback and regulation. This variable is often examined and regulated through the assessment of the process of teaching and learning. The product-oriented examination must be reformed. All the exams for students should be made up according to the aims of teaching.

Besides the cognitive domain, there is the affective domain. "Attitude assessment" can be used for a reference to encourage and arouse the students; to affirm their diligence, success, and achievement, and to guide them to carry out self-education.

Choosing the Method of Assessment

The multiplicity and complexity of factors in teaching result in vagueness in the assessment of teaching. Teachers tried their best to surmount this problem. They analyzed patterns of assessment, and through many kinds of tests and exams, finally developed *synthesized assessment*. Yet, the reliability and the applied value of quantitative assessment have seldom been demonstrated. Some research work has ended only at the elementary stage of the theory and at the lowest level of experiment.

In 1987 we began our research project by classifying assessment into two types in connection with an experiment in schools. One type applies the socalled *fuzzy mathematics* and educational measurement to the assessment of teaching. The other judges the educational targets clearly and directly through investigation and observation and is called *experiential judgement*. The first can obtain more accurate results in fuzzy areas, those suitable for macroscopic assessment. The second can evaluate the processes and the results of teaching activities, and is suitable for microcosmic regulation The purpose of these two types is to simplify and quantify the decision target, and to strengthen experiential judgement so as to overcome the errors and subjective factors in experiential judgement.

In experiential judgement, we should still pay much attention to overcoming the errors related to "time", "process", and "occasion" Because time affects feedback, process affects assessment, and occasion affects the regulation and control of teaching, it is advantageous to set up an integrated and effective system of assessment.

3. APPLIED RESEARCH

Assessment Mechanisms in the Teaching Process

One of the purposes of assessment is to encourage students to study hard and to promote their activity in cognitive areas.

The control of the teaching aim dimension To control the teaching aim dimension means to grasp the structure and quality of the teaching aims. The mathematics syllabus divides learning into four degrees: *learning, understanding, grasping,* and *mastering.* As to knowledge, the following problems should be in focus: "to know or not to know", "to understand or not to understand", "can do or can't do", and "to be or not to be skillful". But how to assess this? We believe that there must be a definite, concrete, and unified standard. Considering our own reality, we divided teaching aims in the cognitive domain into *two kinds* and *four levels.*

The two kinds are the *basic* aim and the *developing* aim. The first is the basic requirements connected to content as determined in the syllabus and in the teaching materials, i.e. the level that most students should reach. The second is a higher requirement in terms of content, an extension of knowledge. This is planned to meet the needs of the students who have high intelligence.

The four levels are *memorizing, understanding, applying,* and *synthesizing.* The levels provide a clear guide for certain forms of study and help us solve such problems as, "What is it?", "How is it?", "How to use it?", and "How to do further research and investigation on it?" For a long time, the problem of ability has been a hot one. People have tried to clear it up. "How to organize training and design classes?" are still problems. Several years of research shows that we should strengthen our teaching process, and focus on the teaching aim we want to reach. In mathematics teaching, we divide ability-related factors into three parts. One is the operation of intelligence and scientific thinking towards the learning requirements of "knowing", "understanding", "doing", and "mastering". The second is the teaching method required for the given content, which includes the distillation of experiences, the identification of methods, the refinement of thought, and means for creative thinking. The third is the generalization of mathematical problems, which trains students to solve problems. This factor embodies the aim of fostering ability; it also embodies the principles of development and activity in teaching. Using these steps we can better solve the problems in mathematics teaching.

Affective education is still very important in teaching. The main points are "interest", "attitude", and "concept". As everybody knows, interest is the motive for study. We should arouse students' *interest*, mobilize all their positive factors, and make them satisfied with what they are learning. *Attitude* means a strict style of study, a practical and realistic one. *Concept*

refers to political education and includes moral and aesthetic standards, as well as a scientific view of the world.

To ensure that teaching aims are put into practice, we set up a network of three levels: an *Estimated Table* for each term, *Assessment Cards*, and *Tests-for-Each-Lesson*. In the Estimated Table for the term, we divide contents into knowledge and ability, define appropriate demands for each level, devise activities for the training of aptitude, and implement this in every unit. The Estimated Table not only relates to contents, but also to the problems we meet in each lesson. Assessment Cards is one of the assessment methods with which we can evaluate our teaching aims. Its purpose is to make clear to teachers and students the teaching structure and the main points, and to direct them to accomplish the teaching aims. Tests-for-Each-Lesson is also a method which helps us to regulate teaching work and ensure the teaching aims are reached.

A system for assessment of cognitive activities To control teaching activities, we set up a system for assessment which is made up of *previous exam*, *process assessment in the class*, *unit assessment*, and *summative assessment*.

"Previous exam" determines whether students are ready to learn the new lesson. Through this assessment we regulate our teaching plan and find some remedial measures for those who fail to understand the previous knowledge. In this way we make all students ready to begin the new lesson.

"Process assessment in the class" is separated into two aspects. One is the internal role played by assessment; the other is feedback in various forms that contain questions and answers, discussion and description, and self-assessment. In order to make full use of the function of process assessment in regulating our work, we make it a basic part of teaching. It contains three steps: intensifying the aim, self-assessment of the students, and the process test. Intensification means to strengthen or clarify the teaching aim, the structure of knowledge, and the level at which every objective in a lesson should be pursued. This step helps the students to assess themselves. Self-assessment promotes the students' self-analysis or self-realization. They can examine themselves according to the lesson, asking: "What have I learned?", "Are there any problems?", "Does the method suit me or not?", "What should I learn next?", etc. So we can see that process assessment serves to determine the teaching results in a class, to locate the problems, and to get to know whether the students have achieved the teaching goal. The time for such an assessment is between 5 minutes and 8 minutes. The assessment should correspond to the teaching plan and the results should be published immediately after the assessment. When students do their own assessment, a *feedback card* must be filled in for the teachers to use for statistics and analysis. This kind of assessment is carried out for every aim; with its function of diagnosis and direction, it serves to ensure that every teaching aim is achieved.

"Unit assessment" consists of a special assessment lesson in a unit, and a test of that unit. The assessment lesson is carried out using Assessment Cards and self-assessment by the students. The teacher gives the necessary guidance. As there are several questions and requirements in the cards, the students may select those that are suitable for themselves. This method is convenient for strengthening the knowledge of students and correcting some of their problems. In order to train the students in self-assessment, and arouse their interest and initiative, a *management map* of study quality should be set up for a term or a unit. This map has proved very convenient for the students, helping them to do their own regulation and self-education.

Assessment in the affective domain. There is no complete system for assessing qualitatively students' study habits and approaches. Some teachers do not know how to deal with this problem. This kind of assessment is not a general judgement of "yes or no", "good or bad", "high or low", "strong or weak", but it has to do with the affective relationships between teachers and the students. Sometimes it can arouse the interest of students, and sometimes it can hinder the interest of students. For the assessment of the student's habits and approaches, we advocate to focus on "rational factors" to promote the interest of the students. We put forward the following principles:

- To stimulate the interest of the students, we start with looking at the teaching method, the materials, the results, and the relationships between the teachers and the students. We try to improve the teaching/learning environment, creating a favorable atmosphere for study.
- To encourage students, we affirm their diligence and effort and speak favorably of their progress and success, so as to help them get greater enjoyment from their own success and satisfaction.
- Affective education cannot be instilled forcibly. The successful way for teachers is to be accessible and open-minded, to help students understand and help each other, and to use vital and rich materials that can fascinate the students. While the students are acquiring knowledge, they are also beginning to understand other things. Their emotions and ideals may, at last, be unified.
- The teachers' passions should affect their students; teachers' own emotions can be used to arouse the interests of the students, so the teachers and the students can understand each other and the students can take an active part in the teaching activities. In practice, two aspects must be stressed: The first is to scrutinize the teaching materials in order to make them more interesting; the second is that the teaching methods must be made more changeable and, they too, more interesting. In a word, *happiness* must be embodied in teaching.

Synthesized assessment in classroom teaching is an important part of educational assessment. In recent years, many plans and methods have come out which are very convenient for research work. Determining the type of assessment, how many elements each contains, how to index the totals, and the kind of statistics to be used, are still problems that need to be solved. To give the assessment of classroom teaching a scientific and practical character, we need to adopt methods that allow for comparisons, measurements, and statistics to be made. We started with an analysis of the basic variables that affect teaching to make our synthesized assessment plan for classroom teaching. Having practiced for a couple of years, we now know more about this problem. This kind of assessment relates to the teachers' teaching goals, and unifies the management of teaching and assessment.

In a word, assessment of the teaching process is for improving and developing teaching, and centers on the achievement of the teaching aims. If we use it in our mathematics teaching, it will bring about appropriate changes in teaching methods and teaching contents.

Analysis of the Result

We achieved good results when we used the type of assessment described above in the mathematics teaching process. It effectively improved the quality of teaching and learning. From our experiment in Fuxin City, started in 1987, results are notable. Through our analysis we can see, in Table 1, that the results continue to improve.

year / mark	1987	1988	1989
Average mark	+15.01	+21.5	+29.9
Pass mark (%)	+27.25	+29.12	+33.0
Excellent (%)	+14.85	+40.0	+26.65

Table 1 *Improvement of teaching results*

Ability also improved as shown in Table 2.

From the two tables, we can see that the teaching experiments have improved the teaching quality in several respects. We computed some statistics on the marks from 3,300 students in 14 schools in Fuxin City. The proportion of passing marks rose from 53.2 percent to 85.7 percent. This indicates that teachers' have improved the quality of their work. All of the teachers now welcome our experiments and our experiments have spread

	Memorize	Understand	Apply	Synthesize
Experimental class	0.92	0.91	0.87	0.76
Control class	0.85	0.83	0.75	0.61

Table 2 *Experimental results*

to many other provinces in our country.

We feel deeply that research on the assessment of teaching must be based on rich educational theory. Assessment is not only for measurement, but also focuses on teaching and education as they appear in natural settings. It can be used to criticize the traditional theory of teaching.

In research on the assessment of teaching, stress must be placed on practicality. The transformation of education will affect the practice of thousands of people. Without teachers, the research will lose its relevance. So it is essential that we bring forth continuously new ideas that combine theory with practice in a dynamical interplay, demonstrating again and again the importance of that link.

On the whole, the purpose of our research is to raise the quality of teaching. We believe we have proven that it can be successful in doing so. The nine-year compulsory education system of our country needs to raise its quality of teaching. We have the responsibility to spread research on assessment to its schools.

Wei Chao-qun & Zhang Hui
Liaoning Education College,
Shen-yang,
China

CHENG ZEMIN & LÜ SHAOZHENG

THE PRACTICE AND STUDY OF EVALUATING MATHEMATICS TEACHING IN CHINA

1. INTRODUCTION AND BACKGROUND

An analysis of ancient educational theories and practical research on the evaluation of mathematics instruction in China is used by the authors to probe into a new way to strengthen mathematics instruction in accordance with China's present reality.

As an investigation has shown, there are key problems in the mathematics instruction of China's middle schools (the three grades students enter after six grades of primary education at ages 12 or 13):

- The qualifications of teachers need enhancing. According to statistics, about 70 percent of teachers are accustomed to the spoon-feeding way of teaching (e.g., chalk and talk).
- The students are poor in rudimentary knowledge. This prevents them from developing their ability and produces great variation in learning outcomes. The pass rate of the students in our city used to be about 5[illegible] percent. A number of students had learned some elementary mathematical terms and symbols, but did not have the competence necessary to analyze and solve problems.
- The students are over-loaded. They are asked to complete "mountains of books and a sea of exercises", which prevent them from making progress at their own initiative. Some students lose interest and motivation to learn and a few drop out of school.
- The solving of problems is restricted by many factors. Such problems are subjectiveness and arbitrariness resulting from a unitary syllabus and teaching materials; general, indistinct instructional objectives, and the conflict between over-loaded students and the overuse of testing.

We have based our research on the theories of a number of educationalists, in ancient or modern times, and take the path of using our cultural heritage, using the experiences of other countries, as well as using references and blazing new trails.

M. Niss (ed.), *Cases of Assessment in Mathematics Education*, 193–201.

2. BASIC IDEAS FOR EVALUATING MATHEMATICS TEACHING

In the traditional theories of education in China, there are a wealth of ideas on instructional assessment which have made a great contribution to enriching and developing international educational theories. "To pass judgement on everything and distinguish evil from good", a principle raised by the great thinker and educationalist Confucius (551–479 B.C.), led the people to follow the "right way" in behavior and judgement. From the book, *On Learning* (403 B.C.), we can see clearly the descriptions of an ancient educational system:

> "Schooling was a formalized educational process which took place in family schools, village schools, city schools and in the national capital university, in which the students were tested systematically every other year. The first year was about their ability to make pauses in reading unpunctuated writings and their motivation to learn; the third year about their attitude to learning and relations to one another; the fifth year their depth and width of learning and respect for teachers; the seventh year their progression in analysis and problem-solving and choice of best friends. Those who passed all the above tests were considered a half success. On the ninth year an all round assessment was put into effect on their reasoning power, their constancy of purpose, and independent thinking. An individual who passed also these tests were a complete success."

The above words provide evidence that, more than 2,000 years ago, instructional objectives and assessment standards in intelligence and aptitudes were highly emphasized.

With regard to the situation in middle schools in China, there seems to be an imperative need for a radical reform of secondary education. However, our experiment of educational assessment is not only an experiment on teaching methodology, but also one on educational models, educational management, and specific ways of assessment.

Obviously, it is an arduous task to meet the needs of economic and social development in regard to the training of qualified personnel. For years the modern educational assessment model has been considered vital to educational progress; the model emphasizes three aspects: systematic consideration of objectives, scientific and practical method, and functional and managerial implementation.

The fundamental principle of the taxonomy of instructional aims are to combine: (1) a holistic approach with the learning hierarchy; (2) scientific method with practicality; (3) developmental testing with the stages of learning. In order to improve the effect and quality of teaching through research, the authors believe one must aim to fully carry out the teaching syllabus, and to make knowledge, competence, consciousness, and action converge for the students by setting up a taxonomy of instruction. We classify the knowledge system in the cognitive domain into four degrees: *simple recall of knowledge*, *comprehension*, *analysis*, and *synthesis*. We classify the affective domain into three categories: *interest*, *attitude*, and *cultivation*.

We put forward the following general slogan: "Teaching with affection, adjusting and controlling in the course, persisting in progressive education, testing and assessing regularly". By so doing, we hope to bring knowledge, competence, consciousness, and practice together, combining educational evaluation with effective teaching.

In view of the above, a research group on educational assessment was constituted of educationalists, research workers, teachers, and administrative staff along the lines of a practice-theory-practice concept. In a process of sampling-surveying summarizing-revising-complementing-perfecting- ting-popularizing they finally obtained some reliable results in their research on educational assessment.

3 ORGANIZATION OF MATHEMATICS EDUCATIONAL ASSESSMENT

After an overall plan was designed to strengthen solidarity among research workers, a one-year plan was drawn up. Four classes of different levels in three middle schools from the 63 middle schools in our city were chosen in 1987 as the first sample; 188 students and several teachers were involved in the research. A continuous dialogue between research personnel and students was promoted through a handbook, *Mathematics Educational Assessment*, that was compiled by our research staff. After a two-year experiment, the handbook was revised at the same time that our research findings were widely disseminated to schools. The handbook consisted of three volumes corresponding to the three grades in (junior) middle school. Owned by every student involved, the handbook was used simultaneously with teaching activities.

The handbook was compiled so as to emphasize global learning and to match students' cognitive development to instructional tasks. Chief among its tasks are the following:

- To classify the levels of knowledge in each chapter of the unified mathematics textbooks.
- To analyze the knowledge structure of mathematical concepts, axioms, theorems, formulas, and laws, etc.
- To sum up the content of each unit into a well defined body of knowledge.
- To change the descriptive demands of the syllabus into four degrees (simple recall of knowledge, comprehension, application, and synthesis) that are visible, measurable and applicable.
- To change abstract teaching aims into definite behavior levels with many examples and exercises as references.
- To make up exercises for formative assessment and schemes for charting an individual's progress.

- To make up reasoning exercises for developing divergent and creative thinking and problem-solving, thus stimulating and keeping alive the individual's motivation to learn and matching talented learners to the teacher's tasks.
- To make up summative exercises and schemes for charting students' achievements during the time; the scheme should deal with figures concerning frequency distributions such as mean, and coefficient of variation.

Under the guidance of their teachers, the students, in their use of the handbook, employ several different approaches to speed up, strengthen, and enhance their learning. Those approaches include the following:

- *From known to unknown.* A pre-view under the guidance of teachers is often a must before learning a new lesson, so that from the beginning students can understand what the objectives of the new lesson are. Diagnostic assessment is carried on during this period.
- *From corrective feedback to reinforcement.* Formative assessment is carried on in the course of learning new material to get more information about the progress of learning and teaching. Each student tracks his progress by listing the number of his correct answers in a particular scheme. In the process, the papers are graded by every student himself (self-assessment), or by other students (peer assessment), or sometimes by the teacher and students together (teacher-centered assessment).
- *From synthesis to consolidation.* Usually in mid-term or at the end of a term summative assessment is carried out. Here, only teachers carry out the schemes for charting students' achievements. Transfer of learning is expected to increase through practice and application.

The students derive their scores after each formative assessment event, according to the following definitions:

$$\text{Individual score} = \frac{\text{Number of correct answers to questions}}{\text{Number of questions}},$$

$$\text{Class score} = \frac{\text{Number of persons passing}}{\text{Nuaber of persons}}.$$

where the failure scores are those below 80%.

The following examples show how instructional objectives can be combined with practice.

Linear Inequalities

1. Simple recall of knowledge

Demands. Know the meaning of the symbols ">", "<", "≥", "≤", be able to read and

write them.

Example Put " > " or " < " between quantities, according to the information given.

(1) a is greater than -2,
(2) b is not greater than 5,
(3) c is not smaller than -6.

2. *Comprehension*

Demands Have a good understanding of the fundamental properties of inequality symbols; be able to use them correctly.

Example If $a < b < 0$, fill in the blanks with proper symbols of inequality.

(1) $(a-b)^2$ __ 0;
(2) $-(a+b)^2$ __ 0;
(3) ab __ 0;
(4) $|a|$ __ $|b|$;
(5) $\frac{1}{a}$ __ $\frac{1}{b}$;
(6) $a-b$ __ 0.

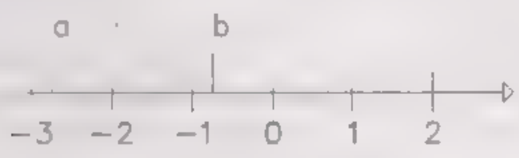
3. *Application*

Demands: Solve problems involving inequality by reasoning.

Example Fill in the blanks with proper symbols of inequality according to the information given.

(1) if $x-5 < -5$ then x __ 0;
(2) if $-3x < 12$ then x __ -4;
(3) if $-\frac{5x}{2} < 0$ then x __ 0;
(4) if $a < b$ and c is not negative, then ac __ bc.

Example:



The number axis above shows the positions of a and b. Which is the correct inequality according to the information given?

(1) $a^2 < b^2$;
(2) $\frac{a}{b} < 1$;
(3) $a < 1-b$;
(4) $\frac{1}{a} < \frac{1}{b}$

Example: If $a > 0$, $b > 0$, $c < 0$, $d > a+b$, compare the values of $ad+bc$ and cd.

Example: If $-2 < x < 2$, and $6x+1 > 7\ 4x$, what is the scope of x in the algebraic expression

$$\frac{6x+1}{7-4x}?$$

4. Synthesis

Demands: Solve the problems creatively by analysis and synthesis, using your skills about inequality.

Example: Suppose

$$x+\frac{2x+7k}{2}-\frac{3x+k}{5}=3-\frac{x-6k}{5}.$$

If x is negative, what is the scope of k?

In order to lead the large-scale experiment effectively, we laid our emphasis mainly on accomplishing the following three tasks:

- *Raising teachers' level of understanding.* Our research group organized varying training programs for teachers more than 40 times, and provided them to about 2,000 trainees. Through mobilizing masses, training core members, enhancing the researchers' quality, developing model teaching, and spreading advanced experience, we disseminated our research findings widely, and our experiment developed quickly and vigorously. Nearly 20,400 students in our city had been involved in the experiment by the end of 1990.
- *Improving teachers' mastery of teaching methodology.* In spite of the different teaching styles of individuals, teaching aiming at the objectives was particularly stressed. In the teaching activities, teachers must always pay special attention to the following conflicts: students' knowledge and the structure of the class, theory and practice, in class or outside activities. Teachers have continuously created suitable new teaching approaches of different styles.
- *Emphasizing teachers' use of assessment.* Educational assessment is present as an important component throughout the whole list of teaching activities. The organic combination of teaching and assessment can surely make a more perfect quality-control system than would be the case if the two were separated.

4. THE ASSESSMENT MODEL OF MATHEMATICS TEACHING

We classify assessment mainly into three modes, *diagnostic*, *formative*, and *summative*. Diagnostic assessment must rely on the results of formative and summative assessment respectively. Formative assessment focuses not only on the prerequisite conditions of the students, but also on the problems that appear in summative assessment in the preceding course, thus providing a basis for remedial teaching. Usually, diagnostic assessment is

carried out together with formative assessment. Summative assessment, aiming at the results of the whole teaching process, can also function as formative assessment. Therefore, there exists an interplay between the three kinds of assessment. Educational assessment is the overlapping effect of diagnostic assessment, formative assessment, and summative assessment. The following Figure can illustrate this relationship:

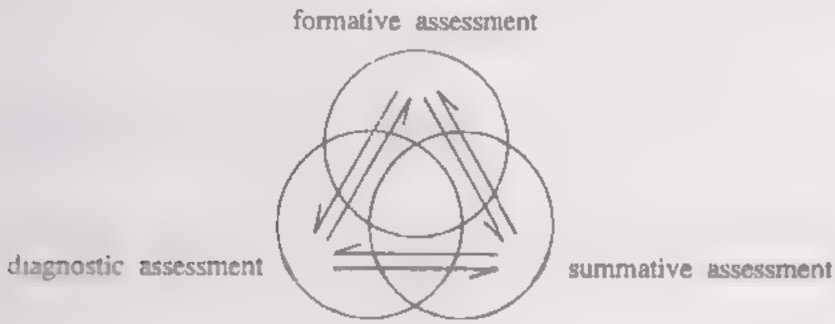


Figure 1

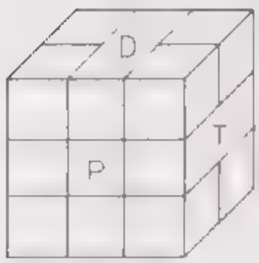
The methods of self-assessment, peer assessment, and teacher-guided assessment are used again and again until, at last, three combinations are realized in a harmonious way. (1) the combination of assessment-in-process and final examinations; (2) the combination of assessment dealing with single items and assessment focusing on comprehensive forms; (3) the combination of description by general words and by specific figures. In this way, our notions and practices of educational assessment develop with apparent and positive effect.

The three dimensional cube below illustrates our educational assessment model.

5. THE EXPERIMENTAL RESULTS OF THE ASSESSMENT MODEL IN MATHEMATICS TEACHING

The research enhanced the quality of teaching. The most conspicuous change in teaching is the move from teaching to examinations to competency-based teaching. Examinations for selection have been changed into proficiency assessments, thus making the teaching suit all the students. The young teachers in our city, through the experiment, have made remarkable progress in mastering the teaching materials and adjusting their teaching methods flexibly, and from them nearly 100 gifted teachers have emerged.

The research has contributed to lightening the overload of students. In the entry test in Huangshi, the students in 1988 and 1990 behaved quite differently; the former did not take part in the experiment, while the latter



P (population)
= (self assessment, peer assessment, teacher-centered assessment)

D (domain)
= (cognitive domain, affective domain)

T (type)
= (diagnostic assessment, formative assessment, summative assessment)

Figure 2

did. The content validity and the difficulty index of the test were equal in two years, and the same type of statistics were used.

Year	Number of students	Excellence rate	Pass rate
1988	5935	11.16%	54.87%
1990	7049	35.64%	71.43%

Table 1 *Comparison of 1988 and 1990*

The figures show a big change in 1990. It is most likely that the experiment on assessment in teaching, in addition to other factors, had a major part in the results.

The research improved the ability of the experiment students. They achieved notable success in contests. In the national middle school mathematics contest of 1990, two students in Huangshi won first-class prizes for the Hubei Province (of which one won the first place by full-marks), one won the second-class prize, five won third-class prizes, and six won the national prizes. In the same year, in the municipal mathematics contest for first year middle school students, 81.4 percent of the winners were from the experimental classes, which included all the first class and second class prizes in the city.

Over 80 percent of students in the experiment reported to like mathematics and took a great interest in the study of mathematics. They regarded the study of mathematics as an arduous, but joyful job. The students in the

experimental classes took an active part in extracurricular activities like sports and games, cultural recreation, and outside reading. They won several national prizes in the competition on inventions and creations.

This indicates that students did not have to do as much homework in order to deal with the examinations. The educational assessment aims to reach the goal of appraising the students of their progress instead of arranging a name list ranked according to their marks. This helps reduce the students' psychological pressure and creates a fine environment for them to develop morally, intellectually, physically, aesthetically, and laboringly.

6. CONCLUSION

As a general result of the research, three changes have come into being. Teachers have moved from mark-domination into objective management; from emphasizing only teaching to both teaching and learning; and from final-examination feedback into a frequent and timely three-dimensional form of feedback.

The experiment on educational assessment is still in an initial stage in China. In order to make the reform really serve the improvement of teaching, we need to study quite a few problems further.

NOTE

Hou Yusheng, Deputy Secretary General of Huangshi Foreign Languages Association of Instruction, translated this article from Chinese.

Cheng Zemin
Huangshi Municipal Education Commission,
China

Lü Shaozheng
Huangshi Teaching and Research Institute,
China

RUTH K. SWEETNAM

ASSESSMENT IN MATHEMATICS WITHIN THE INTERNATIONAL BACCALAUREATE

1. INTRODUCTION

The *International Baccalaureate Organization* (*IBO*) is an international non-governmental organization holding consultative status with UNESCO. Its legal status is that of a Foundation under the supervision of the Swiss Federal Government in accordance with the Swiss Civil Code.

The program of studies leading to the examinations for the *International Baccalaureate Diploma* consists of a two-year period of study for students aged between sixteen and nineteen. It is designed to be comprehensive, demanding and yet within realistic reach of suitable candidates throughout the world. Based on the pattern of no single country, it represents the desire of the founders to provide students of different linguistic, cultural and educational backgrounds with the intellectual, social and critical perspectives necessary for further study and the adult world that lies ahead of them. The Diploma is currently accepted as an entry qualification to university education in over sixty-five different countries.

To put this into perspective in terms of figures the following (Table 1) are those for the May 1990 examination session measured against May 1980. The growth of the *IB* and the increase in popularity is easily demonstrated by this comparison.

	May 90	May 80
Number of schools participating	392	76
Number of candidates examined	13,733	2,845
Number of candidates entered for the Diploma	4,286	1,005
Number of candidates awarded the Diploma	3,188	765
Number of subjects and levels examined	188	150
Number of subject entries examined	37,169	7,968
Total number of nationalities of candidates	157	142
Number of candidates examined in Mathematics	5,741	1,200

Table 1 *IB enrolment*

M. Niss (ed.), Cases of Assessment in Mathematics Education, 203–211.

2. SUMMARY OF THE CURRICULUM AND EXAMINATION

The curriculum consists of six subject Groups:

Group 1	Language A1 (first language) including the study of selections from World Literature.
Group 2	Language B (second language) or a second language A1.
Group 3	Study of Man in Society, History, Geography, Economics, Philosophy, Psychology, Social Anthropology, Organization and Management Studies.
Group 4	Experimental Sciences; Biology, Chemistry, Applied Chemistry, Physics, Physical Science, Environmental System.
Group 5	*Mathematics*; Mathematics, Mathematics with Computing, Mathematical Studies, Mathematics with Further Mathematics.
Group 6	One of the following options: (a) Art/design, Music, Latin, Classical Greek, Computing Studies (b) A school-based syllabus approved by IBO. Alternatively a candidate may offer instead of a Group 6 subject: a third modern language, a second subject from the Study of Man in Society, a second subject from Experimental Sciences.

To be eligible for the award of the Diploma all candidates must

- offer one subject from each of the above Groups;
- offer at least three and not more than four of the six subjects at Higher Level and the others at Subsidiary Level;
- submit an Extended Essay in one of the subjects of the IB curriculum;
- follow a course in the Theory of Knowledge;
- engage in CAS Activities representing Creativity, Action and Service.

Candidates may also offer single subjects, for which they will receive a Certificate.

Examinations from subjects in Groups 3 to 6 may be taken in any one of the three working languages: English, French or Spanish.

Assessment may be external (written or oral) or internal (moderated externally). Each subject is graded from 1 (very poor) to 7 (excellent). Bonus points may be awarded (or penalty points deducted) for both the Extended Essay and Theory of Knowledge. A Diploma is normally awarded to candidates who have scored 24 points or more provided certain conditions are achieved.

3. ROLE OF MATHEMATICS IN THE DIPLOMA

Mathematics holds a unique position within the IB Diploma as it is the only subject which is compulsory. Other Groups offer a variety of subjects but in contrast Group 5 offers programs which are all strictly mathematical in nature.

These separate programs have been carefully created to cater for the wide range of student ability and interest. Each has been designed for a particular group of students.

Mathematics at Higher Level is intended for those who have "good" mathematical ability. Some study the subject because they have a genuine interest in it and also because they enjoy meeting the challenges and problems which it produces, whilst others need mathematics for their future studies in this subject or in other closely related subjects such as physics or engineering.

Mathematics at Subsidiary Level is designed to provide a background of mathematical thought and a reasonable level of technical ability for those not intending to undertake Higher Level. It normally provides a sound mathematical basis for those intending to pursue studies in subjects which have a more limited degree of dependence, e.g., chemistry, biology, economics, etc.

Mathematics with Computing at Subsidiary Level is intended for students with good mathematical ability but, in addition to providing a sound background based on mathematical techniques, it allows the student to gain a working knowledge of programming developed in the context of mathematics. The program has a 55% overlap, in terms of content and assessment, with Mathematics at Subsidiary Level.

Mathematical Studies at Subsidiary Level is intended to provide a realistic mathematics course for students with varied backgrounds and abilities. The skills needed to cope with the mathematical demands of a technological society are developed but no great expertise is required. The intellectual level is comparable to the two previous Subsidiary Level courses.

Further Mathematics at Subsidiary Level may only be taken in conjunction with Mathematics at Higher Level. It is intended for students who plan to specialize in mathematics at university and extends their knowledge of the topics contained in the Higher Level program in addition to the introduction of new areas for study.

4. MODES OF ASSESSMENT

Until 1979, all courses in mathematics were externally assessed by written examination. With the emergence of Mathematical Studies which placed greater emphasis on application, the notion of a different type of assessment was introduced, that of an internally assessed component. This mode

of assessment was subsequently extended to Mathematics with Computing, both courses being designed to develop skills in researching and applying techniques. Other mathematics courses have remained entirely externally assessed by written examination. For all examinations in mathematics, the final mark is calculated by summing the marks obtained from each component of the examination to produce a total mark which is then converted to a grade on the 1 to 7 scale.

External Assessment

Until 1981, all mathematics examinations, except for Further Mathematics, consisted of two papers, one containing multiple-choice questions, the other containing longer, more structured questions.

However, multiple-choice testing was discontinued at this time for the following reasons. Firstly, there was no mechanism in place for pre-testing questions; and secondly, the system made no allowance for method of working, being an "all-or-nothing" form of assessment. Multiple-choice testing was therefore deemed inappropriate and dropped in favor of short answer questions which in addition to testing the breadth of the curriculum are designed to allow students partial credit for correct method of working.

The second paper (and that of Further Mathematics) contains questions designed to test the depth of the curriculum. These require extended responses and sustained reasoning. Marks are allocated on the grounds of method accuracy, and clarity of expression.

Internal Assessment

Mathematical Studies and Mathematics with Computing each have an element of internal assessment which contributes up to 20% of the student's final mark. In both cases the teacher assesses the work of the candidate according to guidelines provided by the IB.

Moderation is carried out by an Examiner based on a submission of sample work. In the case of Mathematical Studies the internal assessment consists of a project containing an extended piece of work developed independently and in the case of Mathematics with Computing it consists of a dossier of selected programs. It is expected that students will be supervised and may receive guidance.

5. DIFFERENTIATION

Differentiation between levels, and in subjects within a level, is achieved by both content and style. Restricting ourselves to the three most popular examinations in mathematics, i.e., Mathematics at Higher Level, Mathematics at Subsidiary Level and Mathematical Studies at Subsidiary Level, this effect can be illustrated though not generalized by the following examples:

Paper 1

Mathematics Higher Level (May 1990, question 12)

12. Given that the equation $4z^3-3z^2+16z-12=0$ has a root $z=2i$, find the other two roots.

(3 marks)

Mathematics Subsidiary Level (May 1990, question 13)

13. Find the values of m for which the quadratic equation

$$x^2+(1-m)x+9=0$$

has two equal roots. **(4 marks)**

Mathematical Studies, Subsidiary Level (May 1990, question 5)

5. A debt was repaid in monthly instalments which formed the geometric progression 1,000, 250,

Calculate the value of the first instalment which was smaller than 2, giving the answer

(i) exactly

(ii) correct to 4 significant figures **(4 marks)**

Paper 2

Mathematical Higher Level (May 1990, question 2)

2. The line l has equation $\frac{x-2}{\ }=\frac{y-3}{2}=\frac{z}{1}$ and the point P has coordinates $(5,7,-3)$. Find the coordinates of Q, the foot of the perpendicular from P onto l and verify that **the length of PQ is $2\sqrt{5}$.** **(7 marks)**

The plane π has equation $\lambda(x+z+2)+\mu(y+2z-3)=0$ where λ and μ are non-zero constants. Show that for all values of λ and μ the plane π contains the line l.

(4 marks)

Hence, or otherwise,

(i) find the equation of the plane which contains l and which passes through the point with coordinates (2,1,0); (4 marks)

(ii) find the equation of the plane which contains l and is perpendicular to the plane with equation $2x+7y-3z=17$. (5 marks)

Mathematics Subsidiary Level (May 1990, question 2)

2. (a) Show that, for all real values of x,

$$(\sin x+\cos x)^2=1+\sin 2x.$$ (3 marks)

(b) Find the values of x in the interval $0\leq x\leq\frac{\pi}{2}$ for which

(i) $\sin x+\cos x=1$;

(ii) $\sin x+\cos x=\sqrt{2}$. (5 marks)

(c) Find the greatest and least values of the expression

$$\sin x+\cos x$$

in the interval $0\leq x\leq\frac{\pi}{2}$. (4 marks)

(d) Find the greatest and least values of the expression

$$\sin^3 x + \cos^3 x$$

in the interval $0 \le x \le \frac{\pi}{2}$. (8 marks)

Mathematical Studies, Subsidiary Level (May 1990, question 3)

3. The temperature, T, of a cup of tea is given as a function of the time, t, after it has been served by

$$T : t \to 15 + 64(2^{-t})$$

T is measured in °C and t is measured in minutes.

(a) Copy and then complete the following table to find values of the temperature, T, at particular times, t. (4 marks)

t	0	1	2	3	4	5
2^{-t}			$\frac{1}{4}$			$\frac{1}{32}$
$64(2^{-t})$			16			2
$T = 15 + 64(2^{-t})$			31			17

(b) Write down the temperature of the tea 4 minutes after serving. (1 mark)

(c) Write down the temperature of the tea when the cup of tea was served. (2 marks)

(d) Draw the graph of T using the table above for $0 \le t \le 5$. (Take 1 cm as the unit for 1 minute on the t-axis and 1 cm as the unit for 10 °C on the T-axis.) (8 marks)

(e) Use the graph to estimate after how many minutes the temperature of the tea is 41°C. (2 marks)

(f) What do you think is the temperature of the room where the tea was served? Give reasons for your estimate. (3 marks)

The marking of examination scripts is carried out by a team of Assistant Examiners from a detailed Markscheme produced by the Chief Examiner. Marks may be awarded for *Method*, *Accuracy* (linked to method), *Correct* answers and clarity of *Reasoning*. These appear as M, A, C or R marks on the Markscheme. Follow Through (FT) marks may sometimes be applied where an incorrect answer fundamentally affects subsequent working.

6. THE GRADE AWARD PROCESS

Only after Chief Examiners have satisfied themselves of the validity of their

Assistant Examiner's marking from a 15% sample of scripts and have decided on the application of moderation factors (if any) can the grade award process begin. It consists of a number of fixed steps as follows:

Nature of the Examination
To gain as wide a view as possible on how the examination was perceived, Chief Examiners read comments submitted by both teachers and Assistant Examiners.

Nature of the Population
Chief Examiners then acquaint themselves with any changes in the nature of the candidate population taking the examination, and satisfy themselves as to the nature of the population in comparison to that of previous years.

Preliminary Information
Mark distributions are provided in graphical form for each element of the examination together with a total mark distribution. These provide an indication of the general overall performance in the examination, but are not used to norm reference the candidate entry.

Grade Boundaries
By considering the work available from candidates, Chief Examiners choose their grade boundaries starting at grades 3 to 4. Having set a provisional boundary, a number of scripts on either side of this boundary are examined and a decision made, based on professional judgement, as to the eventual boundary position. In this way all the grade boundaries are established from Grade 1 (very poor) to Grade 7 (excellent). It should be emphasized that boundaries are chosen on the basis of qualitative judgements. IB does not norm reference its candidate population.

Predicted Grades from Schools
When the grade boundaries have been finalized, a set of CSR data (*Confidential School Reports* containing predicted grades) is provided. This allows Chief Examiners to access the overall CSR versus IB grade differences. However, CSR data do not precipitate moving grade boundaries to fit teacher predictions. It is the Chief Examiner's standard, already decided, which is applied, not that of the teachers.

Where candidates are two or more grades below the predicted grade, scripts may be checked for clerical errors, but they may not be remarked at this point.

7. AWARD OF THE DIPLOMA

In detail, the grading scheme in use for the IB examinations is as follows:

1 — very poor;
2 — poor;
3 — mediocre;
4 — satisfactory;
5 — good;
6 — very good;
7 — excellent.

The diploma is awarded to candidates whose total score, including any bonus or penalty points, reaches or exceeds 24 points and does not contain any failing conditions such as a Grade 1 or Grade 2 at Higher Level; a Grade 1 at Subsidiary Level; more than three Grades 3, etc.

8. FUTURE DEVELOPMENTS

Research is presently being undertaken into the validity of the various aspects of mathematics examinations. In this connection, correlation coefficients have been calculated between the components and the whole as illustrated in the tables below.

Mathematics Higher Level

May 1989 Examination Session

Paper 2	0.8350 *0.8357*	
Total mark	0.9303 *0.9308*	0.9786 *0.9777*
	Paper 1	Paper 2

N = 1280
Key: Pearson Correlation Coefficients
Spearman Correlation Coefficients

Mathematics Subsidary Level

May 1989 Examination Session

Paper 2	0.8066 *0.8110*	
Total mark	0.9126 *0.9190*	0.9710 *0.9753*
	Paper 1	**Paper 2**

N = 2332
Key: Pearson Correlation Coefficients
Spearman Correlation Coefficients

One might expect this high correlation but the reasons for it may be more elusive. Are we testing the same skills in each component, and if so could the extent of examining be reduced, or is there some central mathematical ability being displayed consisting of skills which are essentially linked?

Another issue presently under discussion is the validity of obtaining the final mark by summing the marks achieved for each component. Present thinking favors the concept of *student profiling* in which the final grade is determined from a matrix of all possible combinations. Thus a series of in-built hurdles are created so that candidates who achieved an even spread, demonstrating development of a range of skills, are favored above those candidates who concentrate their energies on a single component, or a small number of skills, and hence obtain an uneven spread.

Ruth K. Sweetnam
International Baccalaureate,
Cardiff,
United Kingdom

INDEX

A

B

C

G

H

I

J

K

L

W

Z

Cases of Assessment in Mathematics Education

An ICMI Study

Edited by
Mogens Niss

This book is probably the first to present a variety of carefully selected cases describing and analyzing in depth and considerable detail, assessment in mathematics education throughout the world. The book is based on work presented at an international ICMI seminar and includes contributions from first-rate scholars from Europe, North America, the Caribbean, Asia and Oceania, and the Middle East.

The cases presented range from thorough reviews of the state of assessment in mathematics education in selected countries, each possessing 'archetypical' characteristics of assessment, to innovative or experimental small- or large-scale assessment initiatives. All the cases presented have been implemented in practice.

The book will be particularly stimulating reading for mathematics educators - at all levels - who are concerned with the innovation of assessment modes in mathematics education.

Kluwer Academic Publishers

ISBN 0-7923-2089-1